

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: March 25, 2005, 23:02:29 ; Search time 931 Seconds

(without alignments)
3930.215 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3238

Sequence: 1 MTMR5AVFRAAAGPAGNPE.....RSRCRLIEQKQPHRTCRK 614

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 5552208 seqs, 297965951 residues

Total number of hits satisfying chosen parameters: 11104416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10054935/runcat.23032005.144821.26181/app_query.fasta.1.775
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rmpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=tblomsum62
-TRANS=human40.csl -LIST=45 -DOCALLIGN=200 -THR SCORE=ext -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=ICL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10054935@cgn_1.1.480_@runcat.23032005.144821.26181
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=6=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELop=6 -DELext=7

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3238	100.0	4372	15	US-10-054-935-1
2	3238	100.0	4372	17	US-10-144-194A-65
3	3238	100.0	4372	18	US-10-491-566-65
4	1448.5	44.7	1529	9	US-09-925-302-212
5	1448.5	44.7	1529	10	US-09-925-302-212
6	1055.5	32.6	615	9	US-09-917-8008-725
7	1053.5	32.5	1280	17	US-10-641-644-322
8	625	19.3	365	10	US-09-918-995-3064
9	381.5	11.8	440	17	US-10-242-535A-31928
10	381.5	11.8	440	17	US-10-085-783A-31928
11	245	7.6	3774	18	US-10-723-860-2538
12	245	7.6	5373	17	US-10-333-314-40
13	245	7.6	5373	18	US-10-737-450-51
14	245	7.6	5438	18	US-10-618-941-2
15	243.5	7.5	9025608	15	US-10-156-761-1
16	238	7.4	2325	18	US-10-723-860-6692
17	234.5	7.2	3107	9	US-09-783-066-5
18	232.5	7.2	5858	19	US-10-488-056-14
19	230.5	7.1	7065	9	US-09-874-923-115
20	230.5	7.1	7065	9	US-09-991-496-115
21	230	7.1	58038	19	US-10-741-600-17942
22	230	7.1	333811	19	US-10-741-600-17681
23	229.5	7.1	2618	18	US-10-476-924-45
24	229	7.1	5373	18	US-10-702-496-5
25	228.5	7.1	3169	15	US-10-037-270-713
26	228.5	7.1	3169	17	US-10-117-722-713
27	227.5	7.0	8459	17	US-10-273-826-3
28	227.5	7.0	8459	17	US-10-274-347-3
29	227.5	7.0	9807	17	US-10-492-380-13
30	227.5	7.0	9807	17	US-10-376-981-13
31	227.5	7.0	9930	17	US-10-210-130-33
32	227.5	7.0	10122	17	US-10-210-130-35
33	227	7.0	6342	10	US-09-814-353-20198
34	226.5	7.0	15952	14	US-10-171-311-51
35	225	6.9	2108	9	US-09-962-832-225
36	225	6.9	2108	17	US-10-240-425-1294
37	224.5	6.9	6227	19	US-10-278-698-100
38	224.5	6.9	6227	19	US-10-278-698-615
39	224.5	6.9	6256	19	US-10-278-698-101
40	224.5	6.9	6256	19	US-10-278-698-616
41	224.5	6.9	16291	15	US-10-017-161-2397
42	224.5	6.9	16291	17	US-10-292-798-2039
43	223.5	6.9	2983	13	US-10-087-192-1517
44	223.5	6.9	3117	9	US-09-752-639-6
45	223.5	6.9	3117	9	US-09-984-198-6

ALIGNMENTS

RESULT 1
US-10-054-935-1
Sequence 1, Application US/10054935
Publication No. US20030143546A1
GENERAL INFORMATION:
APPLICANT: Origene Technologies, Inc
TITLE OF INVENTION: BREAST CANCER TRANSCRIPTION FACTOR GENE AND USES
FILE REFERENCE: 16U 107 RI
CURRENT APPLICATION NUMBER: US/10/054, 935
CURRENT FILING DATE: 2002-01-25
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 4372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (78)..(1922)
OTHER INFORMATION:

US-10-054-935-1

Alignment Scores:

Pred. No.:	1,09e-225	Length:	4372
Score:	3238.00	Matches:	614
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-10-054-935-2 (1-614) X US-10-054-935-1 (1-4372)

```
QY 1 MetThMeArSerAlaValPheLyAlaAlaAlaProAlaGlyYbAnProGlu 20
Db ATGACCATGATGATCCCGGTGTTCAAGCGCGCGCCCTCCCGCGCAATCTTGAG 137
QY 21 GlnArgLeuAspTyrGluArgAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
Db CAGCGACTGACATACAGCGGCTGCGGCTTGGCGGCGCCCAAGACAGACCTTGGGCG 137
QY 41 AlaGluAlaHisPheLeuProArgHisArgLyLeuLySgluProGlyProProLeuAla 60
Db GCCGAGCCCACTTCTCCCGCGCACCGTACGCTCAAGAGCGCGGCGCCCGCTGCGC 257
QY 61 SerSerGlnGlyGlySerProAlaProSerProAlaGlyGlyGlyGlyArgGly 80
Db TCTCTCCAGGCGCGAGCGCCCGCGCTTCCCGCGCGCTGCGCGCAAGCGCGCGCG 317
QY 81 LeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlyGlnGlySerTyrGlySerVal 100
Db TTCTTACTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 377
QY 101 ProLeuProCybProProProAlaThrLySgluAlaGlyTleGlyGlyProAlaAla 120
Db CCTTCGCCCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 437
QY 121 AlaGlyAlaGlyCybSerProArgProLyGlyGlnAlaValLeuProGlnGlnGly 140
Db GCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 497
QY 141 SerLeuValAlaAlaAlaLySgluProThrProThrAlaGlyAlaAspLySgluAlaAla 160
Db TCTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 557
QY 161 SerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeuProGlyPro 180
Db TCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 617
QY 181 ProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlnGlyArgTyrLyS 200
Db CCAACCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 677
QY 201 SerMetArgLySerProLeuGlyGlyGlyGlySerGlyAlaSerSerGlnAlaAla 220
Db AGTATAGGAGAGAGCGCTCTCGGCGGCGTGGCGGCTCGGAGCTTCAATCAAGCGCG 737
QY 221 CybLeuLySgluGlnLeuLeuLeuGlnLeuAspLeuTleGlnGlnGlnGlnGlnLeu 240
Db TGCCTCAACAGATCCCTTCTGCTGCAATTCGACCTCATCGAACGACGACGACGCTG 797
QY 241 GlnAlaLySgluLySgluGlnGlyGlnGlyLeuLySserGluArgAspThrLeuAlaArg 260
Db CAGCGCAAGGAGAAAGGATCGAGAGCTGAAGTCAAGAGAGACACGCTCTTGCCTCGG 857
QY 261 IleGluArgMetGluArgArgMetGlnLeuValLySbAspAsnGlnLySgluArgHis 280
Db ATTGAACGATGAGAAAGCGGATGCACTGTAAAGAGATTAACGAAAGAAAGGACAC 917
QY 281 LySLeuPheGlnGlyTyrGluThrGluGluArgGluGluGluThrGluLeuSerGlyLyS 300
Db AAGCTGTTTCAGGCGCTATGAACTGAAAGAGAGAGAAACAGAGCTATCTGAGAAATTT 977
QY 301 LySLeuGlnCySgluProGluLeuSerGluThrSerGlnThrLeuProProLySProPhe 320
```

```
Db 978 AAACGAGCTGCGAGCGGAGCTTTCGAGACATCCCACTGCTCCGCAAGCCCTTC 1037
QY 321 SerCySgluArgSerGlyLySgluHisLySArgLySserProPheGlySerThrGluArg 340
Db 1038 TCATGTGGCGGAGTGGAGAGGAGACATTAAGAAATCCCATTTGAGAGTACAGAAAG 1097
QY 341 LybThrProValLySgluLeuAlaProGluPheSerLySgluValLySThrProLyS 360
Db 1098 AAGACTCTGTTAAAGAGCTGCTCCTGAATTTTCAAAAGTCAAAACAAAACTCTTAAG 1157
QY 361 HisSerProTleLySgluGluProCySgluSerLeuSerGluThrValCySbAspArg 380
Db 1158 CACTCTCTATTAAAGAGAACCTGTGTCTTATCTGAACCTGTTGTAACGTGA 1217
QY 381 LeuArgSerGlnGluThrProGluLySProArgSerSerValAspThrProProArgLeu 400
Db 1218 TTGAGAGCGCAAGAAACCCCAAGAAAGCCCGGTCTTCACTGAGACCCCAACAGACTC 1277
QY 401 SerThrProGlnLySgluProSerThrHisProLySgluLySAlaPheSerSerGluTle 420
Db 1278 TCCACTCCCAAAAGGAGCCAGACCCATCCCAAGAGAAAGCCCTTCAAGTGAATA 1337
QY 421 GluAspLeuProTyrLeuSerThrThrGlnMetTyrLeuCybArgThrHisGlnProPro 440
Db 1338 GAAGATTCCCGTACTTTCACACAGAAATGATTTGTGTGTGCGACACAGCTCCC 1397
QY 441 ProSerProLeuProLeuArgGluSerSerProLySgluGlnGluThrValAlaArgCyS 460
Db 441 CCAATACCGTTACATTAAGGAAATCTCTCCAAAGAGAGAGACTGATGAGAAAGTGT 1457
QY 461 LeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTyrArgAsp 480
Db 1458 CTGATCCCAATCAAGTGTTCAGAGAAACTTCACTGTGCTGTTCTTCTTGAAGGAG 1517
QY 481 HisSerValGluProLeuArgAspProAsnProSerAspLeuLeuGluLeuAspAsp 500
Db 1518 CACTCAGTAGAGCTCTTAAGGAGCCCAATCTTCAAGACCTTTTGAAGACCTGAGTAG 1577
QY 501 SerValPheSerLySgluAlaLySgluLeuAspGluLySArgArgLySArgTyr 520
Db 1578 AGTGTGTTTTCAGAGCGCATGCAAACTGAGACTGATGAGAGAGAGAGAAAGATGG 1637
QY 521 AspIleGlnArgIleArgGlnGlnArgIleLeuGlnArgLeuGlnLeuArgMetTyrLyS 540
Db 1638 GATATTCAAGAGATCAGGAAACAAAGATTTTACACGCACTGCAAGCTCAAGATGAT 1697
QY 541 LySbLyGlyTleGlnGlnSerGluProGluValThrSerPhePheProGluProAspAsp 560
Db 1698 AAGAAAGGAATTCAGGAATCTGAGCTGAGGTTACTCATTTTCCCTGAGCCAGATGAT 1757
QY 561 ValGluSerLeuMetIleThrProPheLeuProValAlaAlaPheGlyArgProLeuPro 580
Db 1758 GTTGAAGTTGATGATTACCCCTTCTGCTGCTGTTGATGATTTGACATTAATCA 1817
QY 581 LySLeuThrProGlnAsnPheGlnLeuProTyrPheAspGluArgSerArgCySArgLeu 600
Db 1818 AATTAATCTCAAGATTTTGAATTAAGTACCTGAGTGAAGAGCGATGACGATGAT 1877
QY 601 GlnIleGlnLySgluGlnThrProHisArgThrCybArgLyS 614
Db 1878 GAGATCCAGAAAGCAACACTCAACGAGCTGTAGAAA 1919
```

RESULT 2

US-10-144-194A-65

Sequence 65, Application US/10144194A

Publication No. US20030215809A1

GENERAL INFORMATION:

APPLICANT: Origene Technologies Inc

TITLE OF INVENTION: Regulated Breast Cancer Genes

FILE REFERENCE: 3U 103-41

CURRENT APPLICATION NUMBER: US/10/144,194A

CURRENT FILING DATE: 2002-06-12


```

: NUMBER OF SEQ. ID NOS:
: SOFTWARE: Patentin vers
: SEQ ID NO 65
: LENGTH: 4372
: TYPE: DNA
: ORGANISM: Homo sapien
: FEATURE:
: NAME/KEY: CDS.
: LOCATION: (78)..(1922)
US-10-144-194A-65

```

Prod. No.:	1.09e-225	Length:	4372
Score:	3338.00	Matches:	614
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	17	Gaps:	0

US-10-054-935-2 (1-614) X US-10-144-194A-65 (1-4372)

[illegible]

Qy	261	116gluArgMetGluAArgMetGlnLeuValLysLysAspAsnGluLysGluArgHis	280
Db	858	ATTGAACGTATGAAAGGCGGATCGACTCGTTAAAGAAAGATTACGAAGAAAGCAAC	917
Qy	281	LysLeuPheGlnGlyTYRgluThrGluGluArgGluGluThrGluLeuSerGluLysIle	300
Db	918	AACGTGTTTCAGGCGCTATGAACATGAAGAGAGAGAAACAGAGCTATCTGAGAAATTT	977
Qy	301	LysLeuGluLysGlnProGluLeuSerGluThrSerGlnThrIleuProProLysProPhe	320
Db	978	AAACTGGAAGTCCACGCGGAGCTTTCGAGACATCCAGACTCTGCTCCCAAGCCCTTC	1033
Qy	321	SerCysGlyArgSerGlyLysGlyVHIsLysArgLysSerProPheGlySerThrGluArg	340
Db	1038	TCATGTGGGGCGAGTGGAAAGGAGCATTAAGAAATCCCATTTTGGAAAGTACGAAGA	1099
Qy	341	LysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLys	360
Db	1098	AAGACTCTCTGTAAAGAGCTGGCTCTCGAATTTTCAAAAGTCAAAACAAAATCTCTAAG	1155
Qy	361	HisSerProIleLysGluGluProCysGlySerLeuSerGluThrValCysLysArgGlu	380
Db	1158	CACCTCTCTATTAAAGAGAACCTGTGTCTTATGTGAACCTGTTGTGAACGTGAA	1211
Qy	381	LeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProProArgLeu	400
Db	1218	TTGAGAGACCAAGAAACCCACAGAAAGCCCGGTCTTCAGTGGAGAACCCACCAAGACTC	1277
Qy	401	SerThrProGluLysGlyProSerSerThrHisProLysGlyLysAlaPheSerSerGluIle	420
Db	1278	TCCACTCCCCAAAGGAGACCCAGACCCATCCAAAGGAGAAAGCTTCTCAAGTGAAGATA	1333
Qy	421	GluAspLeuProTYRLeuSerThrThrGluMetTYRLeuCysArgTYRHisGlnProPro	440
Db	1338	GAAATTTGCCGTACCTTTCCACACAGAAATGATTGTGTCTGTGGCACACAGCTCCC	1399
Qy	441	ProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrValAlaArgCys	460
Db	1398	CCATCACCGGTATCATTTACGGGAATCTCTCCAAAGAGAGAGAGACTGTGACAGAGTGT	1455
Qy	461	LeuMetProSerSerValAlaGluGluThrSerValLeuAlaValProSerThrArgAsp	480
Db	1458	CTGATGCCATAGGTGTGACGAGAAACCTTCAAGCTTGGCTCTTCTTTGGAGGAGAC	1511
Qy	481	HisSerValGluProLeuArgAspProAsnProSerAspLeuLeuGluAsnLeuAspAsp	500
Db	1518	CACTCAGTAGAGCTCTTAAGGAGCCCAATCTTCACAGCTTTTGGAGAACCTGTGATGAC	1577
Qy	501	SerValPheSerLysArgHisAlaLysLeuGluLeuAspGluLysArgArgLysArgTrp	520
Db	1578	AGTGTGTTTTTCGAAGCGGACATGCAAAATGGAGCTGATGAGAAAGAGAAAGATGG	1633
Qy	521	AspIleGlnArgIleArgGluGluArgIleLeuGlnArgLeuGlnLeuArgMetTYRlys	540
Db	1638	GATTTTCAAGAGATCAAGGAAACAAAGATTTTACAGCACTGCGACTCAGAAATGATTAA	1699
Qy	541	LysLysGlyIleGlnGluSerGluProGluValThrSerPhePheProGluProAspAsp	560
Db	1698	AAGAAAGGAATTCAGGAATCTGAGCCTGAGGTTCCTCATTTTTCCTGAGCCAAATGAT	1755
Qy	561	ValGluSerLeuMetIleThrProPheLeuProValAlaPheGlyArgProLeuPro	580
Db	1758	GTTAAAGATTGATGATTAACCCCTCTCTGCTGTGTGAGCAATTGGACGACATTTACCA	1811
Qy	581	LysLeuThrProGluAsnPheGluLeuProTrpLeuAspGluArgSerArgCysArgLeu	600
Db	1818	AAATTAATCTCACAGAAATTTTGAAGCTACCTGTGTTGATGAGCGGATGCGATGCAGATTG	1877
Qy	601	GluIleGlnLysLysGluThrProHisArgThrCysArgLys 614	
Db	1878	GAGATCCAGAAAGAACCACTCTCACCGAGCGTGTAGAGAA 1919	

RESULT 3

US-10-491-566-65
 ; Sequence 65, Application US/10491566
 ; Publication No. US20040257197
 ; GENERAL INFORMATION:
 ; APPLICANT: Origène Technologies Inc
 ; TITLE OF INVENTION: Regulated Breast Cancer Genes
 ; FILE REFERENCE: 3U 103 R1
 ; CURRENT APPLICATION NUMBER: US/10/491, 566
 ; NUMBER OF SEQ ID NOS: 148
 ; SOFTWARE: Patent version 3.1
 ; SEQ ID NO 65
 ; LENGTH: 4372
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (78)..(1922)
 ; OTHER INFORMATION:
 ; US-10-491-566-65

Alignment Scores:

Pred. No.:	1.09e-225	Length:	4372
Score:	3238.00	Matches:	614
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	18	Gaps:	0

US-10-054-935-2 (1-614) x US-10-491-566-65 (1-4372)

QY 1 MetThMeArGSeRAlaValPheLyAlaAlaAlaProAlaGlyGlyAsnProGlu 20
 Db 78 ATGACCATGATGATCCCGGTTCGAAGGCGCGCGCCCTCCCGGCAATCTGTAG 137
 QY 21 GlnArgLeuAspTyrGluArgAlaAlaLeuGlyGlyProGluAspGluProGluAla 40
 Db 138 CAGCGATGACTACAGAGCGGCTGCGCGCTGCGCGCGCCAGAGCGAGCTGGGCG 197
 QY 41 AlAGlAlaHisPheLeuProArGHisArgLyLeuLyGlyGluProGlyProLeuAla 60
 Db 198 GCGGAGCCACCTCTCCCGCGGCAACCTCAAGTCAAGAGCGCGGCGCCCGCTGCGC 257
 QY 61 SerSerGlnGlyGlySerProAlaProSerProAlaGlyGlyGlyGlyGlyGly 80
 Db 258 TCTCTCCAGGCGCGAGCCCGCGCTTCCCGCGCGCTGCGCGCAAGCGCGCGG 317
 QY 81 LeuLeuLeuProAlaGlyAlaAlaProGlyGlyGlyGlyGlyGlyGlyGlyGly 100
 Db 318 TTGTTACTCCCG 377
 QY 101 ProLeuProCysProProProAlaThrLyGlnAlaGlyYllleGlyGlyGluProAlaAla 120
 Db 378 CCTTCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 437
 QY 121 AlAGlYAlAGlYcysSerProArGProLyGlyGlnAlaValLeuProIleGlnThrGly 140
 Db 438 GCGGAGCG 497
 QY 141 SerLeuValAlaAlaAlaLyGluProThrProTTPAlaGlyAspLyGlyGlyAlaAla 160
 Db 498 TCTCTGCG 557
 QY 161 SerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuProProGlyPro 180
 Db 558 TCCCGCGCTGCCACCGCTCGGACCGCGCGGAGCCCGCACCACTACTTCCCGGCGG 617
 QY 181 ProProLeuAlaProThrAlaThrAlaGlyYThrLeuAlaAlaSerGlnGlyArgTTPlys 200
 Db 618 CCACCCCTCGAG 677
 QY 201 SerMetArgLySerProLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 220

Db 678 AGTATGAGAGAGAGCCCTCTCGGGGTGTGTGGCGGCTCCGAGCCTCCAGTCAAGCGCGC 737
 QY 221 CysLeuLyGlnIleLeuLeuLeuGlnLeuAspLeuIleGlnGlnGlnGlnGlnLeu 240
 Db 738 TCCCTCAACAGATCCTTCTGCTGCAATTGACCTCATGCAAGAGAGAGAGAGAGAG 797
 QY 241 GlnAlaLyGlnGlyGlnIleGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 260
 Db 798 CAGGCGGAG 857
 QY 261 IleGluArgMetGluArgArgMetGlnLeuValLyLyAspAsnGlyGlyGlyGlyGly 280
 Db 858 ATTGAACGATGAGAAAGCGGATGACGCTGTTAAAGAAAGATACAGAGAAAGAGCAG 917
 QY 281 LyLeuPheGlnGlyTyrGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGly 300
 Db 918 AAGCTGTTTCAAGGCTATGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 977
 QY 301 LyLeuGluCysGlnProGluLeuSerGluThrSerGlnThrLeuProProLySProPhe 320
 Db 978 AAATGAGATGCGAGCGGAGCTTCCGAGACATCCAGACTCTGCTCCAGAGCCCTTC 1037
 QY 321 SerCysGlyArgSerGlyLySgLyHisLySArgLySerProPheGlySerThrGluArg 340
 Db 1038 TCATGTGGCGGAGTGTGAAGAGAGACATTAAGAGAAATCCCATTTGGAAGTACAGAAAG 1097
 QY 341 LyThrProValLyLyLeuAlaProGluPheSerLySValLyThrLySProLyS 360
 Db 1098 AAGACTCCTGTTAAAGCTGCTCCTGAATTTTCAAAAGTCAAAACAAACCTCTTAAG 1157
 QY 361 HisSerProIleGlyGlnGluProCysGlySerLeuSerGlnThrValLySArgGly 380
 Db 1158 CACTCTCTATTAAAGAGAACCTGTGTCTTATTCGAAGCTGTTGTAACCTGAA 1217
 QY 381 LeuArgSerGlnGluThrProGluLySProArgSerSerValAspThrProProArgLeu 400
 Db 1218 TTGAGAGAGCCAGAAAGCCCAAGAAAGCCCGGCTTCACTGAGACACCCACCAAGACTC 1277
 QY 401 SerThrProGlnLySgLyProSerThrHisProLyGlyGlnAlaPheSerSerGluIle 420
 Db 1278 TCCACTCCCAAGATA 1337
 QY 421 GluAspLeuProTyrLeuSerThrThrGluMetTyrLeuCysArgTTPHisGlnProPro 440
 Db 1338 GAAGATTTCGCGTTCATCCACACAGAAAGTATTTGTGCTTGAGAGACAGCCTCCC 1397
 QY 441 ProSerProLeuProLeuArgGlySerSerProLySgLyGlnGlnThrValAlaArgCys 460
 Db 1398 CACTACCGTTACATTACGGAATCTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGTGT 1457
 QY 461 LeuMetProSerSerValAlaGlyGlnThrSerValLeuAlaValProSerTTPArgAsp 480
 Db 1458 CTGATGCATCAAGTGTTCAGAGAAAGCTTCAAGTGTGCTGCTCTTCTTGAAGAGAG 1517
 QY 481 HisSerValGluProLeuArgAspProAsnProSerAspLeuLeuGlnLeuAspAsp 500
 Db 1518 CACTCAGTAGAGCTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1577
 QY 501 SerValPheSerLySArgHisAlaLySLeuGlnLeuAspGlnLySArgArgLySArgTTP 520
 Db 1578 AGTGTTTTCGAAGGCGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1637
 QY 521 AspIleGlnArgIleArgGlnGlnArgIleLeuGlnArgGlnLeuArgMetTyrLyS 540
 Db 1638 GATATTCAAGAGATCAGGAGAAAGAAATTTTAACGCGAGCTGAGCTCAGAAATGATTA 1697
 QY 541 LyLySgLyIleGlnGlnSerGluProGluValThrSerPhePheProGluProAspAsp 560
 Db 1698 AAGAAAGAGATTCAGAAATCAGAGCTGAGGTTACTCTTTTCCCTGAGCAGAGATGAT 1757
 QY 561 ValGluSerLeuMetIleThrProPheLeuProValAlaPheGlyArgProLeuPro 580
 Db 1758 GTTGAAGATTGATGATTAACCCCTTGTGCTGTGTGAGCATTTGAGAGACCATTAACA 1817

Oy		581	LysLeuThrProGlnAsnPhgIleuLeuProTrpIleuAapGIuArgSerPArGCysArgIeu	600
Db		1818	AAATTAACCTCCAGCAATTGTGACTACCCTGGTTTGATAGCGGACGACGATTC	1877
Oy		601	GluIleGlnLysLysGlnThrProHisArgThrCysArgLys	614
Db		1878	GAGATCCAGAAAGAACCAACCTCACCGGACGTGTAGAAA	1919
	RESULT 4			
	US-09-925-302-212			
	; Sequence 212, Application US/09925302			
	; Patent No. US2002004941A1			
	GENERAL INFORMATION:			
	; APPLICANT: Rosen et al.			
	; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies			
	; FILE REFERENCE: PA104			
	; CURRENT APPLICATION NUMBER: US/09/925,302			
	; CURRENT FILING DATE: 2001-08-10			
	; PRIOR APPLICATION NUMBER: PCT/US00/05918			
	; PRIOR FILING DATE: 2000-03-08			
	; PRIOR APPLICATION NUMBER: 60/124,270			
	; PRIOR FILING DATE: 1999-03-12			
	; NUMBER OF SEQ ID NOS: 896			
	; SOFTWARE: PatentIn Ver. 2.0			
	; SEQ ID NO 212			
	; LENGTH: 1529			
	; TYPE: DNA			
	; ORGANISM: Homo sapiens			
	US-09-925-302-212			
	Alignment Scores:			
	Pred. NO.:	4.09e-96	Length:	1529
	Score:	1448.50	Matches:	284
	Percent Similarity:	95.64%	Conservative:	1
	Best Local Similarity:	95.30%	Mismatches:	6
	Query Match:	44.73%	Indels:	7
	Dbl:	9	Gaps:	2
	US-10-054-935-2 (1-614) x US-09-925-302-212 (1-1529)			
Oy		181	ProbProleuAlaProThrAlaThrAlaGlyThrIleuAlaAlaSerGIuGIuYArgTrpIys	200
Db		3	CCACCCCCTCGGCCCCACGCCGCCGCGAGACCTTGGCGGCCAGGAGGCGAGTGGAG	62
Oy		201	SerMetArgLysSerProLeuGIuYGIuYGIuYSerGIuYAlaSerSerGIuAlaAla	220
Db		63	AGTATGAGGAAGAAGCCCTCTCGGGGGTGTGTGGCGGCTCGGAGCCTCCAGTAGGCGGCC	122
Oy		221	CysLeuLysGlnIleLeuLeuLeuLeuAspLeuIleGlnGlnGlnGlnGlnLeu	240
Db		123	TGCCCAAACAAGATCTTCTGCTGCATTGGACCTCATCGAACAACAGACGACGAGCTG	182
Oy		241	GlnAlaLysGIuLysGlnIleGlnGIuLeuLysSerGIuArgPThrIleuLeuAlaArg	260
Db		183	CAGGCGCAAGAAAAAGAGATCGAGGAGCTGAAGTACAGAGAGAGACACGCTCTCTCGG	242
Oy		261	IleGIuArgMetGIuArgArgMetGlnIleuValLysLysASPasnGIuLysGIuArgHis	280
Db		243	AITGACCTATAGAAAGGCGGATGACGCTGTAAAGAAAGAAAGAAAGAAAGGCAAC	302
Oy		281	LysLeuPhgIleGlnLYrGIuThrGIuGIuArgGIuGIuThrGIuLeuSerGIuLysIle	300
Db		303	AAGCTGTTTCAAGGCTTAGAAACTGAAGAGAGAGAAACAAGACTATCTGAGAAAATT	362
Oy		301	LysLeuGIuCYrGlnProGIuLeuSerGIuThrSerGIuThrLeuProProLysProPhe	320
Db		363	AAACTGGAGTGCACACCGAGCTTTCCGAGACATCCCAAGCTCTCTCCCAAGCCCTTC	422
Oy		321	SerCYrGIuArgSerGIuLysGIuHisLysLysArgLysSerProPhgGIuSerThrGIuArg	340
Db		423	TCAATGGGCGGATGAGAAAGGACATTAAGAAATATCCCATTTTGGAGATCAAGAAACA	482

Oy	341	lysrhrProValylslyLeuAlaProGluPheSerLysValIythrThrlysrhrProLys	360
Db	483	AAGACTCTGTTAAAAGAACTGGCTCTGAATTTTAAAAGTCAAAACAAAACCTCTAAG	542
Oy	361	HisserProIleLysGluGluProCysGlySerLeuSerGluThrValCysIlysaRgIu	380
Db	543	CACCTCTCTAATTAAAGAGAAACCTGTGTGTTCTTATCTGAACACTGTTGTAAAGTGA	602
Oy	381	LeuIlyrSerGluGluThrProGluLysProAlrGserSerValAspThrProProAlrGleu	400
Db	603	TTGAGAGACCAAGAAACCCCAAGAAAGCCCGGCTCTCAGTGACACCCCAACAGACTC	662
Oy	401	SerThrProGluLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIle	420
Db	663	TTCACTCCCAAGAGGAGCCACGACCCATCCCAAGAGAAAGCCCTTCCAAGTGAATA	722
Oy	421	GluAspLeuProTyrLysSerThrThrGluMetTyrLeuCyArgTPhisGlnProPro	440
Db	723	GAAATTTGGCCGTACCTTTCCACACACAAATATGATTTGTGTGCTTGGACACAGCTCC	782
Oy	441	ProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrValAlaArgCys	460
Db	783	CCATCACCGTATACCATTTACGGGAATCTCTCCAAAGAAAGAGAGAGACTGTAGCA-----	836
Oy	461	LeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerThr	478
Db	837	-----AGTAAGGCATATAGAAACACTTGCTCTT--ATACCTTAGTGG	875
RESULT 5			
US-09-925-302-212			
/ Sequence 212, Application US/09925302			
/ Publication No. US20030064072A9			
/ GENERAL INFORMATION:			
/ APPLICANT: Rosen et al.			
/ TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies			
/ FILE REFERENCE: PA104			
/ CURRENT APPLICATION NUMBER: US/09/925,302			
/ PRIOR FILING DATE: 2001-08-10			
/ PRIOR APPLICATION NUMBER: PCT/US00/05918			
/ PRIOR FILING DATE: 2000-03-08			
/ PRIOR APPLICATION NUMBER: 60/124,270			
/ PRIOR FILING DATE: 1999-03-12			
/ NUMBER OF SEQ ID NOS: 896			
/ SOFTWARE: PatentIn Ver. 2.0			
/ SEQ ID NO 212			
/ LENGTH: 1529			
/ TYPE: DNA			
/ ORGANISM: Homo sapiens			
US-09-925-302-212			
Alignment Scores:			
Pred. No.:	4,09e-96	Length:	1529
Score:	1448.50	Matches:	284
Percent Similarity:	95.64%	Conservative:	1
Best Local Similarity:	95.30%	Mismatches:	6
Query Match:	44.73%	Indels:	7
DB:	10	Gaps:	2
US-10-054-935-2 (1-614) x US-09-925-302-212 (1-1529)			
Oy	181	ProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGluIyargTpyLys	200
Db	3	CCACCCCTCGCGCCACCGCCACCGCCCGGACCTTGCGCGCCAGCGAGGCGAGATGGAAG	62
Oy	201	SerMetArgLysSerProLeuGlyGlyGlyGlyGlySerGlyValAspSerSerGlnAlaAla	220
Db	63	AGTATGAGGAAGAAGCCCTCTCGGGGGTGTGTGGCGGCTCGAGGACCTCCAGTCAAGGCGGCC	122
Oy	221	CysLeuLysGlnIleLeuLeuLeuGlnLeuAspLeuIleGluGlnGlnGlnGlnLeu	240
Db	123	TGCCTCAAAACAGATCTTCTGTGTGAATTTGACCTCATGTGAACAGAGAGACGACGAGCTG	182
Oy	241	GlnAlaLysGluLysGluIleGluGluLeuLysSerGluArgAspThrLeuLeuAlaArg	260

Db	183	CAGGCCAAGAAAAAGAGATTCAGAGAGCTGMAAGTCCAGAGAGACACGCTCTTGGCTCGG	242
Qy	261	llegluarqmetgluargargmetgineuvalylslaspaanglulysgluarghis	280
Db	243	ATTGAACGTATGAAAGGCGGATGCGCTGTGTAAAGAGATTAACGAGAAAGGAC	302
Qy	281	lysleuphegnlgltyrtrglulthrugluuargglugluuthrclulneuserglulyslle	300
Db	303	AAGCTGTTCCAGGGCTATGTAAACTGAAGAGAGAGAGAAACAAGGTATCTGAGAAATTT	362
Qy	301	lysleuglucysglinproglulneuserglulthrserglulthrleuproprolysprophe	320
Db	363	AAACTGAGTCCACGCGGAGCTTTCCGAGACATCCCAAGCTTGGCTCCAGAGCCCTTC	422
Qy	321	sercysgluargserglulysglulhislysaarglysseproheglisethertrguarg	340
Db	423	TCATGTGGGCGGAGTGGAAAGGAGCATTAAGAAATTCCTCATTTGGAAGTACAGAAAG	482
Qy	341	lysthrprovalyllyslaleuualaprogluipheaserlyslvalylsthrlysthrprolys	360
Db	483	AAGACTCTCTGTAAAAAGCTGGCTCCTGAATTTTCAAAAGTCAAAACAAACCTCCTAAG	542
Qy	361	hisserproilleyaglugluuprocysgluyserserleuserglulthrvalcyslvsarqglu	380
Db	543	CACCTTCTCTATTAAAGAGGAAACCTGTGGTCTCTTATCTGAACGTGTTGTAAACGTGAA	602
Qy	381	leuargserglinlulthrproglulysproargserservalaspthrproproargleu	400
Db	603	TTGAAGAGACAGAAACCACCAAAAGCCCGGTCTTCAGTGGACACCCACCAAGACTC	662
Qy	401	serthrproglulnlysglyproserthrhisprolyglulysalaphesergerglulle	420
Db	663	TCCACTCCCCAAAGGAGACCACGACCCCATCCCAAGAGAAACCTTCTCAAGTGAGATA	722
Qy	421	gluaspluuprotylenseerthrtrnglumetylencysargtrphisginpropro	440
Db	723	GAAAGATTGCGGTACTTTCACCAACAGAAATGTATTGTGTGTGGACACAGCCCTCC	782
Qy	441	proserproleuproleuarggluuserprolysleugluulthrvalalargcys	460
Db	783	CCATCACCGTTTACMTTACGGGAATCTCTCCAAABAGAGAGACTGTAACA-----	836
Qy	461	leuwetprosersevalalaglglulthrservalleuualprosertrp	478
Db	837	-----AGTAAAGCATATGAGAACACTTGTCTT---ATACCTTAGTGG	875
RESULT 6			
US-09-917-800A-725			
Sequence 725, Application US/09917800A			
Patent No. US20020119462A1			
GENERAL INFORMATION:			
APPLICANT: Mendrick, Donna			
APPLICANT: Porter, Mark			
APPLICANT: Johnson, Kory			
APPLICANT: Caslele, Arthur			
APPLICANT: Elashoff, Michael			
APPLICANT: Gene Logic, Inc.			
TITLE OF INVENTION: Molecular Toxicology Modeling			
FILE REFERENCE: 44921-5038-US			
CURRENT FILING DATE: 2001-07-31			
PRIOR APPLICATION NUMBER: US 60/222,040			
PRIOR FILING DATE: 2000-07-31			
PRIOR APPLICATION NUMBER: US 60/222,880			
PRIOR FILING DATE: 2000-11-02			
PRIOR APPLICATION NUMBER: US 60/290,029			
PRIOR FILING DATE: 2001-05-11			
PRIOR APPLICATION NUMBER: US 60/290,645			
PRIOR FILING DATE: 2001-05-15			
PRIOR APPLICATION NUMBER: US 60/292,336			
PRIOR FILING DATE: 2001-05-22			
PRIOR APPLICATION NUMBER: US 60/295,798			

	PRIOR FILING DATE:	2001-06-06	
	PRIOR APPLICATION NUMBER:	US 60/297,457	
	PRIOR FILING DATE:	2001-06-13	
	PRIOR APPLICATION NUMBER:	US 60/298,884	
	PRIOR FILING DATE:	2001-06-19	
	PRIOR APPLICATION NUMBER:	US 60/303,459	
	PRIOR FILING DATE:	2001-07-09	
	NUMBER OF SEQ ID NOS:	1740	
	SOFTWARE:	Patentin Ver. 2.1	
	SEQ ID NO 725		
/	LENGTH:	615	
/	TYPE:	DNA	
/	ORGANISM:	Rattus norvegicus	
/	FEATURE:		
/	OTHER INFORMATION:	Genbank Accession No. US20020119462A1 A1102093	
/	NAME/KEY:	misc.feature	
/	LOCATION:	(1)..(615)	
/	OTHER INFORMATION:	n = a or c or g or t	
US-09-917-800A-725			
 Alignment Scores:			
Pred. No.:	5,85e-68	Length:	615
Score:	1055.50	Matches:	197
Percent Similarity:	97.55%	Conservative:	2
Best Local Similarity:	96.57%	Mismatches:	4
Query Match:	32.60%	Indels:	1
DB:	9	Gaps:	1
 US-10-054-935-2 (1-614) x US-09-917-800A-725 (1-615)			
QY	13	AAlAPrOAlAGlYglYvsnProGluglnAgLueAsPryGluArgAlaAlaLeuGly	32
Db	3	GCCCCCTGGCCCGGCACACCCTCAGAGCACTGTGAATCGAACGGGGCTGGGCTGGGCC	62
QY	33	GLPYrGlunApSjUpProGlyAlaAlaIagluAlahIsPhenLeuProArghIsArgLYsLeu	52
Db	63	GGGGCCGAGGACGAAgTCCGGGGGGCGCCGAAgCCCACCTTCCTCCCGCATCTAAgCTC	12
QY	53	LysGluProGIyPrObProleuAlasErSerGInglySerProAlaProSerProla	72
Db	123	AMGAGAGCCGGGCCCCCGCTGGCTTACTCCAGGGGGAGGCCCGCGCTTCAGCT	18
QY	73	GLYCyE---GLYGLYLySGlYArGGLYLEuLeuLeuProAlAGLYAlaLaProGlyGln	91
Db	183	GGCTGGGGCGCGGCAAGGGCCGGGGTTTGTTACTCCGGCGGGGGCCCCCGGGCAG	24
QY	92	GLnGluJnsErTrPGlYglYSerValProLeuProCYsProProProAlathrlYsgIn	11
Db	243	CAGGAGAAGAGCTGGGGCGGTTGGTGGCTTCCTGCTGTCGCCGCCCGCTACCAACA	30
QY	112	AlAGlYlleglYglYupProAlaAlaIaglyAlaGLYCySeSProArGProlYsTYr	13
Db	303	GCGGGCATCGGGGGGAGCGAGTCGAGCGGGCGGTGGCTGAGAGCCCCCGGCCAAGTA	36
QY	132	GLnaIvalleuProllleglnThrGLYserLeuValAlaAlaLYsGluProThrPro	15
Db	363	CAGCGGTGTGTCCTTGCATTGAGAGGGCTCTCTGTGGCGCGGCCCAAGAGCTACGCC	42
QY	152	TyrAlAGlYsPlYsGLYglYAlaAlaaserProAlaAlaThrcLasErspProAlagly	17
Db	423	TGGCTGGGGACAAAGGTGGGGGTGCCCTCCCCAGCTGCCACGGCTCGAGCCGGGGGA	48
QY	172	ProProleuProleuProglYProProProleuAlaProThzAlathrAlaglyThr	19
Db	483	CCCCCACCTACTCTTGGCCGGGGCCACCCCTGGCCCAAGCCACTGCGGGACC	54
QY	192	IeuAlaAlaSerJnglYArGTPLYsSerMeArGLYsSerProleuGlyglYglY	21
Db	543	CTGGCGGCGCAGTAGGGGAGATGAGAGAGATTAAGAAAGAGCCCTCTCGGGGGTGGCGG	60
QY	212	GLYserGLYAla 215	
Db	603	NGCTGGGAGCC 614	

RESULT 7
US-10-641-643-322
Sequence 322, Application US/10641643
Publication No. US20040077003A1
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
Susan G. Stuart
Jeffrey J. Sellhammer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
GENE EXPRESSION
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/641,643
FILING DATE: 14-Aug-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 322:
SEQUENCE CHARACTERISTICS:
LENGTH: 1290 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TLYMNOR01
CLONE: 140704
SEQUENCE DESCRIPTION: SEQ ID NO: 322 :
US-10-641-643-322
Alignment Scores:
Pred. No.: 1.74e-67 Length: 1290
Score: 1053.50 Matches: 207
Percent Similarity: 93.69% Conservative: 1
Best Local Similarity: 93.24% Mismatches: 6
Query Match: 32.54% Indels: 8
DB: 17 Gaps: 2
US-10-054-935-2 (1-614) x US-10-641-643-322 (1-1290)
QY 258 LeuAlaArgTLeGluArgMetGluArgArGMeGlnuValLysLysAspAsnGluLys 277
Db 3 CTTCCTCGATTGAACGATGAAAGCGAGTCACTGTAAAGAAAGATTAACGAGAAA 62
QY 278 GluArgHisLysLeuPheGlnGlyTyrGluThrGluGluArgGluGluThrGluLeuSer 297
Db 63 GAAAGGACAAGCTGTTTCAGGCGCTATGAAAGTGAAGAGAGAGAGAAACAGAGCTATCT 122
QY 298 GluLysLysLeuGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 317
Db 123 GAGAAATTAATTAATGAGTGGAGTCCAGCCGAGAGCTTTCGAGACATCCAGACTGCTCC 182
QY 318 LysProPheSerCysGlyArgSer-GlyLysGlyHisLysArgLysSerProPheGlySe 337

Db 183 AAGCCCTTCTCAGTGGCGGAGTGGAGAAAGGACATAAAGGAAATCCCATTTGGAG 242
QY 337 rThrGluArgLysrThrProValLysLysLeuLapProGluPheSerLysValLysThrLy 357
Db 243 TACAGAAAGAAAGACTCTGTTTAAAGAGCTGGCTCTGTAATTTTCAAAAGTCAAAACAA 302
QY 357 sThrProLysHisSerProLysGluGluProCysGlySerLeuSerGluThrValCy 377
Db 303 AACTCTAAGACACTCTCTATTAAAGAGAACCTGTGGTCTTATCTGAAACTGTTTG 362
QY 377 sLysArgGluLeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrPr 397
Db 363 TAAACGTGAATTGAGAGGACCAAGAAACCCAGAAAAGCCCCGTCTTCAGTGACACCCC 422
QY 397 oProArgLysSerThrProGluLysGlyProSerThrHisProLysGluLysLapPheSe 417
Db 423 ACCAAGACTCTCCACTCCCAAAAGGAGACCCAGACCCATCCCAAGAGAAAGCTTCTC 462
QY 417 rSerGluLysLeuAspLeuProTyrLeuSerThrThrGluMetTyrLeuCysArgTrpHi 437
Db 483 AAGTGAGATAGAAAGATTGGCGGTACCTTCCACACAGAAATGATTTGTGCTGGCA 542
QY 437 sGlnProProProSerProLeuProLeuArgLysSerProLysLysGluGluThrVa 457
Db 543 CCAGCTCCCGCCATCCGTTACATTAACGGAGATCTCTCCAAAGAGAGGAGACTGT 602
QY 457 lAlaArgCysLeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProse 477
Db 603 AGCA-----AGTAAGCATAGAGAACCTTGCTCTT--ATACCTTA 641
QY 477 rTrp 478
Db 642 GTGG 645
RESULT 8
US-09-918-995-30064
Sequence 30064, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30064
LENGTH: 365
TYPE: DNA
ORGANISM: Homo sapiens
US-09-918-995-30064
Alignment Scores:
Pred. No.: 6.6e-37 Length: 365
Score: 625.00 Matches: 118
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.16% Mismatches: 0
Query Match: 19.30% Indels: 0
DB: 10 Gaps: 0
US-10-054-935-2 (1-614) x US-09-918-995-30064 (1-365)
QY 281 LysLeuPheGlnGlyTyrGluThrGluGluArgGluGluThrGluLeuSerGluLysIle 300
Db 8 GAGCTTTTCAGGCGCTATGAAACTGAAGAGAGAGAGAAACAGACTATCTGAAATTT 67
QY 301 LysLeuGluCysGlnProGluLeuSerGluThrSerGlnThrLeuProPolysPProPhe 320
Db 68 AAACGTGAGTGCAGCGGAGCTTTCGAGACATCCAGACTCTGCTCCAGACCTTC 127

```
QY 321 SerCyGlyArgSerGlyIysGlyHisIleArgLysSerProPheGlySerThrGluArg 340
DB 128 TCATGTGGGGCGGAGTGGAAAGGAGACATAAAGAAATCCCATTTTGAAGTACAGAAA 187
QY 341 LysThrProValIleLysIleuAlaProGluPheSerIleValIleThrLysThrProLys 360
DB 188 AAGACCTCCTGTTAAAGAGTGGTCTCTGATTTTCAAAAGTCAAAACAAAACCTCTTAG 247
QY 361 HisSerProIleLysGluGluProCyGlySerIleuSerGluThrValCysIleArgGlu 380
DB 248 CACTCTCTATTAAGAGAACCCCTGTGGTCTCTTATCTGAACCTGTTGTAAACGTGAA 307
QY 381 LeuArgSerGluGluThrProGluLysProArgSerValAspThrProProArg 399
DB 308 TTGAGAGAGCCAAAGAACCCCGAGTCTTCAGTGGACACCCACCAAGCA 364

RESULT 9
US-10-242-535A-31928
; Sequence 31928, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31928
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4)..(4)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-31928

Alignment Scores:
Pred. No.: 3,96e-19 Length: 440
Score: 381.50 Matches: 76
Percent Similarity: 85.56% Conservative: 1
Best Local Similarity: 84.44% Mismatches: 6
Query Match: 11.78% Indels: 7
DB: 17 Gaps: 2

US-10-054-935-2 (1-614) x US-10-242-535A-31928 (1-440)
QY 389 LysProArgSerSerValAspThrProProArgLysSerThrProGluLysGlyProSer 408
DB 7 AAGCCCCGGTCTTCAATGAGACACCCCAACAGACTCTCCATCCCAAAAGGAGACCCAGC 66
QY 409 ThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuProTyrlleuSerThr 428
DB 67 ACCCATCCCAAGAGAAAGCCCTTCTCAAGTAGAGATTAAGATTTGCCGTATCTTTCCACC 126
QY 429 ThrGluMetTyrlleuCyArgTrpHisGlnProProProSerProLeuProLeuArgGlu 448
DB 127 ACAGAAATGATTGTGTCTGTGGACCAAGCTCCCAATCACCGTTACCATTAACGGGA 186
QY 449 SerSerProLysArgGluGluThrValAlaArgCysLeuMetProSerSerValAlaGly 468
DB 187 TCCTCTCCAAAGAGAGAGAGACTGTAGCA-----AGTAAGGCATAG 228
```

```
QY 469 GluThrSerValLeuAlaValProSerTrp 478
DB 229 AGAACACTTGCTCTT---ATACCTTAGTG 255

RESULT 10
US-10-085-783A-31928
; Sequence 31928, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31928
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4)..(4)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-31928

Alignment Scores:
Pred. No.: 3,96e-19 Length: 440
Score: 381.50 Matches: 76
Percent Similarity: 85.56% Conservative: 1
Best Local Similarity: 84.44% Mismatches: 6
Query Match: 11.78% Indels: 7
DB: 17 Gaps: 2

US-10-054-935-2 (1-614) x US-10-085-783A-31928 (1-440)
QY 389 LysProArgSerSerValAspThrProProArgLysSerThrProGluLysGlyProSer 408
DB 7 AAGCCCCGGTCTTCAATGAGACACCCCAACAGACTCTCCATCCCAAAAGGAGACCCAGC 66
QY 409 ThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuProTyrlleuSerThr 428
DB 67 ACCCATCCCAAGAGAAAGCCCTTCTCAAGTAGAGATTAAGATTTGCCGTATCTTTCCACC 126
QY 429 ThrGluMetTyrlleuCyArgTrpHisGlnProProProSerProLeuProLeuArgGlu 448
DB 127 ACAGAAATGATTGTGTCTGTGGACCAAGCTCCCAATCACCGTTACCATTAACGGGA 186
QY 449 SerSerProLysArgGluGluThrValAlaArgCysLeuMetProSerSerValAlaGly 468
DB 187 TCCTCTCCAAAGAGAGAGAGACTGTAGCA-----AGTAAGGCATAG 228

RESULT 11
US-10-723-860-2538
; Sequence 2538, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnick, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
```

;; TITLE OF INVENTION: Methode for Screening for Soft Tissue Sarcoma Modulators
;; FILE REFERENCE: 05882.0193.NPUS01
;; CURRENT APPLICATION NUMBER: US/10/723,860
;; PRIORITY FILING DATE: 2003-11-26
;; PRIOR APPLICATION NUMBER: 60/429,739
;; PRIORITY FILING DATE: 2002-11-26
;; NUMBER OF SEQ ID NOS: 8393
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 2538
;; LENGTH: 3774
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-723-860-2538

Alignment Scores:
Pred. No.: 2,49e-08 Length: 3774
Score: 246.00 Matches: 181
Percent Similarity: 31.53% Conservative: 81
Best Local Similarity: 21.78% Mismatches: 291
Query Match: 7.60% Indels: 278
Gaps: 32

US-10-054-935-2 (1-614) x US-10-723-860-2538 (1-3774)

QY 11 AlaAlaAlaProAlaGlyGlyAsnProGluGlnArgLeuAspTyrGluArgAlaAla 30
Db 393 GCTGCGGACAGACAGCGCGG-----CCGCTCGCGTATGGCGGAGCTGTACGA 440
QY 31 LeuGlyGlyProGluAsp---GluProGlyAlaAlaGluAlaHisPheLeuProArgHis 49
Db 441 GCGCGAGGTCCGCGAGATGCGCGCGCGGCTGCTGCGCTGG-----CGCGGC 488
QY 50 ArgLeuLeuGlyGluPro-----GlyProProLeuAlaSerSerGlnGlyLeuPro 67
Db 489 GCGCGGTACAGTACGCTTGGAGACAGACACCTCTGAGAGACATCCGCGACGTGCCCA 548
QY 68 AlaPro-----SerProAlaGlyCysGlyGlyLeuGlyArgGlyLeuLeuPro 84
Db 549 GCGCTTACAGACAGAGCGCGCGGAGAGAGCGGCGCGCGCGCGCGCGCTGGC 608
QY 85 Ala-----GlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 97
Db 609 GCGCTTCCGCGAGAGCGCGAGCGCGCGCGCGCTGAGCTGAGAGAGAGCGCGCGCT 668
QY 98 ---GlySerValProLeuProCysProProProAlaThrLysGlnAlaGlyLeuGly 116
Db 669 GCGAGAGAGTGGCGCTACCTCGCGCGCGCACCA-----701
QY 117 GluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTyrGlnAlaValLeuPro 136
Db 702 ---CCAGGAAGAGGTGGCGGAGCTGCTCGCGCAGATCCAGGCGTCCGG-----746
QY 137 IleGlnThrGlySerLeuValAlaAlaAlaLysGlnProThrProTPrAlaGlyAspLys 156
Db 747 ---CGCGCGCGAGCGAGATGCGAGCGCGAGACCGCGAGCGCTTGAAGTGCAGCTGAC 803
QY 157 GlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuPro 176
Db 804 GTCGCGCTCGCGAGATTCGCGCGAGCTTGAAGGCGACGCGGTGACAGACAGCTGCA 863
QY 177 LeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlu 196
Db 864 GTCGAGAGA-----GTGGTTCGAGTGAAGGCTGAGACCGACTGTCGAGAGGAGCA 914
QY 197 GlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGly-----211
Db 915 GGT-----GAACAAGACGCTATGCGCTCAGCGCAGAGAGAGATTAAGTACGCG 965
QY 211 -----211
Db 966 GCGTACAGCTGAGCGCGAGACACAGAGCTGGAGGCACTGAAAAACCAAGCACTACT 1025
QY 212 -----GlySerGlyAlaSerSerGlnAlaAlaCysLeu-LysGlnI 225

Db 1026 GGAGAGCAGCGCTGTGAGCTGAGAGACCGCTCATCAGGCGCGACATTGCTTCAACAGGA 1085
QY 225 eLeuLeuLeuGlnLeuAspLeu---IleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 244
Db 1086 AGCCATTCAAGCTGAGCTGAGCGCTGAGTGAAGAACACCAAGTGGAGATGGCCCGCCAGCT 1145
QY 244 uLysGlnIleGlnGlnLeu-----250
Db 1146 GCGAGATACAGAGACTGCTCAATGTCAGATGCTGTGATATAGATAGCCGCTTA 1205
QY 250 -----250
Db 1206 CAGAAAACTCTGGAAGTGAAGAGTGTGCGATTTGGCTTGGCCCAATTCCTTCTGCT 1265
QY 250 -----250
Db 1266 TCAGAGAGACTCCCAAAATTCCTCTGTGTCTCACTCACATTAAGGTGAAAAAGCAGGA 1325
QY 251 -----LysSerGluArgAspThrLeuAlaArgIleGluArgMe 264
Db 1326 GAAGATCAAGTGTGAGAGAGTGTGAGAGAGAACTGTGATT-----GTGAGAGACA 1379
QY 264 tGluArgArgMeCglnLeuVal-----LysLysAspAsnGlu 277
Db 1380 GACAGAGAGACCCCAAGTGACTGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1439
QY 277 sGluArgHisLeuLeuPheGlnGlyTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 292
Db 1440 GAGAGAGCGCAAG 1499
QY 293 -----GluThrGluLeuSerGlnLys 300
Db 1500 AACAAAGTCCCGCCAGCAG 1559
QY 300 eLysLeuGlnCysGln-----305
Db 1560 AAG 1619
QY 305 -----305
Db 1620 CCGAGCGGAAGTCAAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1679
QY 306 -----ProGluLeuSerGlnThrSerGlnThrLeuProProLysPr 319
Db 1680 ACCGCTGAG 1739
QY 319 oPheSerCysGlyArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrG 339
Db 1740 CCGCAG 1790
QY 339 uArgLysThrProValLysLeuAlaProGluPheSerLysValLysThrLysThrPr 359
Db 1791 GCGCAAGTCTCCAG 1850
QY 359 oLysHisSerProIleLysGlnGln-----ProCysG 370
Db 1851 GCGCAAGTCCCGAGTGAAG 1910
QY 370 ySerLeuSerGlnThrValCysLysArgGlnLeuArgSerGlnGln-----385
Db 1911 GCGCAAGTCTCCAAAG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1964
QY 386 -----ThrProGluLysProArgSerSe 393
Db 1965 TGAAGAGCCAAAGTCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2024
QY 393 rVal-----AspThrProProArgLysSerThrProGlnLysGlyProSerTh 409
Db 2025 AGTGAAG 2084
QY 409 rHisPro-----LysGlnLysAlaPheSerSerGlnIleGluAs 422

QY 208 GlyGlyGlyGlyGlySerGlyAlaSerSerGlnAlaAlaCysLeuYsglnIleLeu 227
Db 1786 GGGATGGAGCCCTGAGGGCTGGAGGCTCAGAGAGGCCCACTGAGAGAGAGGCGGCC 1845
QY 228 LeuGlnLeuAspLeuIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 247
Db 1846 CTCGAG 1905
QY 248 GlnGlnLeuLeuSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 267
Db 1906 TGGCAGCTGAG 1959
QY 268 MetGlnLeuValLeuYsglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 287
Db 1960 GCAGAGCTGGCCAG 2004
QY 288 ThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 295
Db 2005 GAGCGGCGGAG 2058
QY 296 GAGCGGCGGAG 301
Db 2059 TGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2118
QY 302 LeuGlnCysGlnProGlnLeuSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 321
Db 2119 GAGAGAGTGAAGCTGAG 2178
QY 322 CysGlnYsglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 334
Db 2179 ATGAGAGCCCTGAG 2238
QY 335 PheGlnSerThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 346
Db 2239 CAGGAGCTGAG 2298
QY 347 GAG 2358
Db 2299 CTCGAG 2358
QY 357 LysThrProLysHisSerProLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 372
Db 2359 GAGCTGGGAG 2418
QY 373 GAGCTGGGAG 2478
Db 2419 TCCTTCGAG 2478
QY 388 GAGCTGGGAG 388
Db 2479 AGGCAATGAG 2538
QY 389 GAGCTGGGAG 399
Db 2539 GGTGTTCGAG 2598
QY 400 GAGCTGGGAG 407
Db 2599 TGGGAG 2658
QY 408 GAGCTGGGAG 421
Db 2659 CACACGCTGGGAG 2692
QY 421 GAGCTGGGAG 436
Db 2693 GAGCTGGGAG 442
QY 437 GAGCTGGGAG 442
Db 2746 GGTGTTCGAG 2805
QY 442 rProLeuProLeuYsglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 456

Db 2806 CCCGTCCTGAG 2865
QY 456 rVal-----AlaArgCysLeuMetProSerSerValAlaGlnGlnGlnGlnGln 471
Db 2866 ACTGCTTGAAG 2925
QY 471 rValLeuAlaValPro 476
Db 2926 CGCGTGTTCGTCGCC 2941
RESULT 13
US-10-737-450-51
Sequence 51, Application US/10737450
Publication No. US20040235071A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc
APPLICANT: Lightcap, Eric S.
APPLICANT: Ecsedy, Jeffrey A.
APPLICANT: Hunter, John Joseph
APPLICANT: Macbeth, Kyle J.
APPLICANT: Tighe Nestor, Michelle
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING
TITLE OF INVENTION: CANCER USING 15986, 2188, 20743, 9148, 9151, 9791, 44252,
TITLE OF INVENTION: 14184, 42461, 8204, 7970, 25552, 21657, 26492, 2411, 15088,
TITLE OF INVENTION: 1905, 28899, 63380, 33935, 10480, 12686, 25501, 17694,
TITLE OF INVENTION: 15701, 53062, 49908, 21612, 38949, 6216, 46863, 9235, 2201,
TITLE OF INVENTION: 6985, 9883, 12328, 18057, 21617, 39228, 49928, 54476, 62113,
TITLE OF INVENTION: 64316, 12264, 32362, 58198, 2887, 3205, 8557, 9600, 9693,
TITLE OF INVENTION: 44867, 53058, 55556, 57658, 2208, 10252, 10302, 14218,
TITLE OF INVENTION: 33877, 10317, 10485, 25964, 14815, 1363, 1397, 14827, 21708,
TITLE OF INVENTION: 38017, 64698, 2179 OR 13249
FILE REFERENCE: MP102-207P1RNMNM
CURRENT APPLICATION NUMBER: US/10/737, 450
CURRENT FILING DATE: 2003-12-16
PRIOR APPLICATION NUMBER: US 60/435,108
PRIOR FILING DATE: 2002-12-20
PRIOR APPLICATION NUMBER: US 60/436,443
PRIOR FILING DATE: 2002-12-23
PRIOR APPLICATION NUMBER: US 60/438,498
PRIOR FILING DATE: 2003-01-07
PRIOR APPLICATION NUMBER: US 60/444,370
PRIOR FILING DATE: 2003-01-31
PRIOR APPLICATION NUMBER: US 60/446,031
PRIOR FILING DATE: 2003-02-06
PRIOR APPLICATION NUMBER: US 60/453,635
PRIOR FILING DATE: 2003-03-11
PRIOR APPLICATION NUMBER: US 60/457,199
PRIOR FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: US 60/462,458
PRIOR FILING DATE: 2003-04-10
PRIOR APPLICATION NUMBER: US 60/466,732
PRIOR FILING DATE: 2003-04-30
PRIOR APPLICATION NUMBER: US 60/469,184
PRIOR FILING DATE: 2003-05-08
Remainder Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 144
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 51
LENGTH: 5373
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(4719)
US-10-737-450-51
Alignment Scores:
Pred. No.: 4 22e-08 Length: 5373
Score: 245.00 Matches: 139
Percent Similarity: 33.07% Conservative: 68
Best Local Similarity: 22.20% Mismatches: 199
Query Match: 7.57% Indels: 222


```
RESULT 15
US-10-156-761-1/C
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 0.000106 Length: 9025608
Score: 243.50 Matches: 132
Percent Similarity: 28.45% Conservative: 37
Best Local Similarity: 22.22% Mismatches: 214
Query Match: 7.52% Indels: 213
DB: Gaps: 19

US-10-054-935-2 (1-614) x US-10-156-761-1 (1-9025608)
QY 4 ArgSerAlaValAlaPheLysAlaAlaAlaAlaProAlaGlyLysAn-ProGluGlnArgLe 23
DB 2781266 AGATCTCCCGCGGACCGCACTTCGCGAAGCCCGCGGGAGAAAGCTCCGCTCTGCCCC 2781207
QY 23 uAerTYrGluArgAlaAlaAlaLeuGlyGlyProGluuAspGluProGlyAlaAlaGluAl 43
DB 2781206 GGGTGCACGGCGAACA-----GGACCGCGGATCGCGGACATCCTCCTCTGG 2781159
QY 43 aHisPheLeuProArgHisArgLysLeuLysGluProGlyProProLeuAlaSerSerg1 63
DB 2781158 AGAACGGCGAACCCTGTCACCGAACTCCTCGCGCTCCACTCGCCCGCGCGAGCGCGCG 2781099
QY 63 nGlyLysSerProAlaArgHisArgLysLeuLysGluProGlyProProLeuAlaSerSerg1 83
DB 2781098 TGGGCGCGCGCGCGCTCTCTACCGCGCGGCGAGTGGGACGGCGGCG----- 2781053
QY 83 uProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 103
DB 2781052 -CCTGCGCGGCGCGCGCGCGCGGGA-CGT-----CGCGCGCC 2781019
QY 103 cCySPProProProAlaThrLysGlnAlaGlyLysGlyGlyGluProAlaAlaAlaGlyAl 123
DB 2781018 GCACCTCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCGCGCGCGCACTC 2780959
QY 123 aGlyCysSerProArgProLysTYrGlnAlaValLeuProLleGlnThrGlySerLeuVa 143
DB 2780958 CTCCTCTCTACCGCGCGCGG-----GCCGCGCGCGCGCGCTCGCGCGCGAGCGCGGC 2780905
QY 143 lAlaAla-----AlaLysGluPr 149
DB 2780904 CTCGCGGTACCGCGCGGTGAGCGATCCACCGGCTCAAGCGCGCGCTCGAGAGACCGAAC 2780845
```

```
QY 149 oThrProTPAlaGlyAspLysGlyGlyAlaAlaAlaSerProAlaAlaThrAlaSerAspPr 169
DB 2780844 GTACCCGACGCTCTCCCGCGCTCTGGCGCGCA-CGCTACCGAGCTCTCCGCGCTTGGCGG 2780786
QY 169 oAlaGlyProProProLeuProLeuProGly----- 179
DB 2780785 CAGCGGCAAGGGCGCATGTCGCCCGGACCGGATGCGCGCGGAGTGCAGAGACAATACCGC 2780726
QY 180 -----ProPro----- 181
DB 2780725 CGCGTCTGCGCGCTGTGACCGCGCTGACCGCGCGCGGACAGCCCGTACCGCGCGCTG 2780666
QY 182 -----ProLeuAlaProThrAlaThrAlaAlaGlyThrLeuAlaAlaSerGluLysAr 198
DB 2780665 TGGCTGTGACCGAGGGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 2780606
QY 198 gTRPlySerMetArgLysSerProLeuGlyGlyGlyGlyGlySerg1AlaSerSerg1 218
DB 2780605 CTCGCGCG-----CGCGCACCGTGTGGGGCTTGGCGCGGCTCTGCTCAGCGATAC 2780555
QY 218 nAlaAlaCysLeuLysGlnLleLeuLeuLeuGlnLeuAspLeuLleGluGlnGlnGln 238
DB 2780554 CCGCGCG----- 2780548
QY 238 nGlnLeuGlnAlaLysGluLysGluLleGluGlnLeuLysSerg1LysArgAspThrLeu 258
DB 2780548 ----- 2780548
QY 258 uAlaArgLleGluArgMetGluArgArgMetGlnLeuValLysLysAspAsnGluLysG1 278
DB 2780548 ----- 2780548
QY 278 uArgHisLysLeuPheGlnGlyTYrGluThrGluGluArgGluThrGluLeuSerg1 298
DB 2780547 -----ACCGGCGCACCGCTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2780516
QY 298 uLysLleLysLeuGluCys-----GlnProGlnLeuSerg1Lys 312
DB 2780515 AGCGACCTGCGCGCGCTGCTCGACGAGTGGAGAGCGGAGCGCGCGCGGTGATGCCAGATC 2780456
QY 312 rGlnThrLeuProProProLysProPheSerCysGlyArgSer----- 325
DB 2780455 GCCCACCGCGCGCGCGCGCGCGCTATGTCGCCGCTCTGCGCGGTACCGCACCGCTCACCC 2780396
QY 326 -----GlyLysGlyHisLysArgLysSe 333
DB 2780395 TGGAGCGGGGCTACGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2780336
QY 333 rProPheGlySerThrGluArgLysThrProValLys-----LeuAlaProG1 350
DB 2780335 GTGGCGCGCGCGGACCGGCGCGCGCGCGGACGCGGACGAGTCCAGGTCCGCGTGGCTCCGTC 2780276
QY 350 uPheSerLysValLysThrLysThrProLysHisSerPro----- 363
DB 2780275 GTGCTACCGAGACGACACCGCGCTACCGCGCTGGAACAACCGAGCGCGCGAGAGAG 2780216
QY 364 -lLysGlnGluProCysGlySerLeuSerg1LysThrAla-----CysAlaArgGluLeuAr 382
DB 2780215 GAGGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAC 2780156
QY 382 gSerGlnGluThrProGluLysProArgSerSerValAspThrPro----- 397
DB 2780155 GGGTCCGAGATCGCGCACCGGCGAGCGAGGTCCTGACGCGCACCGCGCGGACCTGCGCTCC 2780096
QY 398 -----ProArgLeuSerThrProGlnLysGlyProSerThrAr 410
DB 2780095 AGCTTACCGTCCCGCGCGCTCGCTGCTGCTGCGCGCGCGACCGCGAACCCCACTACGGC 2780036
QY 410 sProLysGluLysAlaPheSerSerg1LleGluAspLeuProTYrLysSerThrThrG1 430
DB 2780035 GCCCGGACATTCGGCGCTGACGAGAGCGGAGCGCGCTGCGACCGCGCTCTCGAACCGG 2779976
QY 430 uMetTYrLeuCybArgTP-----HisGlnProPro----- 440
```

```

Db 2779975 GACACGAGGTGTGGGTGACCTGCCCCGAGGCGGCGGACGACCGCTCCGCGCGCG 2779916
QY 441 ---ProSerProLeuPProLeuAArgGluSerSerProLyseLysGluGluThrValAlaAr 459
Db 2779915 CAGGCCGGGACCGACCGCCCCCGGACATCCGCC----- 2779882
QY 459 gCy8LeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrpAr 479
Db 2779881 -----CCGACCGCACCTACCTGTCACCGGCGGCTCTCGGCGGCTCGGCTG-- 2779834
QY 479 gaSPHisSerValGluPProLeuAArgAspProAsnProSer 492
Db 2779833 -----TCACCGCCCGCGAAGCTCTGCGCCCTCG 2779807

```

Search completed: March 26, 2005, 04:30:50
 Job time : 5329 secs

This Page Blank (uspto)

Best Local Similarity 97.7%; Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2796 GGGGAGAGTGTGGAATAGTGTCTTCTTGTGCTGGCAAAATGTACATCTTGAACAA 2855
DB 1 GGGGGAGAGTGTGGAATAGTGTCTTCTTGTGCTGGCAAAATGTACATCTTGAACAA 60
QY 2856 ACAGAGTACCTTAATAGCTTCCATCTCACTTTGTAATAATTTGTAATGTGTCCAT 2915
DB 61 ACAGAGTACCTTAATAGCTTCCATCTCACTTTGTAATAATTTGTAATGTGTCCAT 120
QY 2916 CTGTGCTCTCTCCCTCCCTGTTTGTAAATATACAGATAGCAGCTCCAGCCACTTTG 2975
DB 121 CTGTGCTCTCTCCCTCCCTGTTTGTAAATATACAGATAGCAGCTCCAGCCACTTTG 180
QY 2976 GTCTCAGTGTAGATCCCTATTAATCTGAAAGAAATAGAGCCAGACCTGTGTG 3035
DB 181 GTCTCAGTGTAGATCCCTATTAATCTGAAAGAAATAGAGCCAGACCTGTGTG 240
QY 3036 TCAGATATATAGAAATTTGCTTTCTTAAGTCTTCAAGATATTTGTGAAACAAAGTAG 3095
DB 241 TCAGATATATAGAAATTTGCTTTCTTAAGTCTTCAAGATATTTGTGAAACAAAGTAG 300
QY 3096 GGTCTATCTCTCAGAGAGTGGGCTTTTATCTTAAAGAAATATGTCCAGATTTAT 3155
DB 301 GGTCTATCTCTCAGAGAGTGGGCTTTTATCTTAAAGAAATATGTCCAGATTTAT 360
QY 3156 TACACTTTTGAAGAGAACCAAGATATGTAGGGTGTGTGGCTGGCCATCAGTGAAGC 3215
DB 361 TACACTTTTGAAGAGAACCAAGATATGTAGGGTGTGTGGCTGGCCATCAGTGAAGC 420
QY 3216 ACAGAGAGAAATGGATACATTTGTGGAGAGAGAAAGAAAGTTCTCCAGAGGCTCCCA 3275
DB 421 ACAGAGAGAAATGGATACATTTGTGGAGAGAGAAAGAAAGTTCTCCAGAGGCTCCCA 480
QY 3276 CTGCTAAGTCTTGTGAGATGTGATCTGTGCTTCTGATTTGACTTTAAAGGAAT 3335
DB 481 CTGCTAAGTCTTGTGAGATGTGATCTGTGCTTCTGATTTGACTTTAAAGGAAT 540
QY 3336 TATTTGGGAGACATGTAGTATTTTGATGATCTTGTCTGCTCTTAATTTCTCTTTTGT 3395
DB 541 TATTTGGGAGACATGTAGTATTTTGATGATCTTGTCTGCTCTTAATTTCTCTTTNNN 600
QY 3396 GT 3455
DB 601 NNN 660
QY 3456 GTGGGCTCTCTATTAAGGAACTGTGTAACTTCACTTGAAGAGATGTAGAGAGAA 3515
DB 661 GTGGGCTCTCTATTAAGGAACTGTGTAACTTCACTTGAAGAGATGTAGAGAGAA 720
QY 3516 TAGAGCTTAATTCACAGAGGGCTCTCATCTCACACCTTAAGAGAGATTTCTAGAAA 3575
DB 721 TAGAGCTTAATTCACAGAGGGCTCTCATCTCACACCTTAAGAGAGATTTCTAGAAA 780
QY 3576 ACTGGGCGAGATTTTCTTTGTTCTCATCATTTTAATGTGCGAGCTGTCAAGTTTCTT 3635
DB 781 ACTGGGCGAGATTTTCTTTGTTCTCATCATTTTAATGTGCGAGCTGTCAAGTTTCTT 840
QY 3636 ACTCTTAACATGATATTTCTTCTGTAAGTGTCCAAAAAGAAAGAACCAATCAGT 3695
DB 841 ACTCTTAACATGATATTTCTTCTGTAAGTGTCCAAAAAGAAAGAACCAATCAGT 900
QY 3696 GTCTCTTGACTTTGTTTGTGATCCCTCAGTTTCTTGTGATTTCAAGATGTGTGGTT 3755
DB 901 GTCTCTTGACTTTGTTTGTGATCCCTCAGTTTCTTGTGATTTCAAGATGTGTGGTT 960
QY 3756 CCTAATTTTGGATAGATGAGAAATTAACATGTGTGTGTGCTTACCCAGGGGAC 3815
DB 961 CCTAATTTTGGATAGATGAGAAATTAACATGTGTGTGTGCTTACCCAGGGGAC 1020
QY 3816 TCCCGAGTTTCTGACTTGAAGTAGACTGAGAGAAATCCAGAGGTGTACTGTGGCCAGT 3875

DB 1021 TCCCGAGTTTCTGACTTGAAGTAGACTGAGAGAAATCCAGAGGTGTACTGTGCCAGAT 1080
QY 3876 TTAAGTAAATTTATTTCTTGTGTTCTCCTCTCCCTGAGGACCTTTATTTTATGTGTC 3935
DB 1081 TTAAGTAAATTTATTTCTTGTGTTCTCCTCTCCCTGAGGACCTTTATTTTATGTGTC 1140
QY 3936 CCTCTTCTAAGTAAATTTCTTGTGTTGATTTGATTTGAGAGAGAGTGTGACAGTAGA 3995
DB 1141 CCTCTTCTAAGTAAATTTCTTGTGTTGATTTGATTTGAGAGAGAGTGTGACAGTAGA 1200
QY 3996 TTAGCAAGTCTCAAGTGCAAAATTAACAGTGTGTAGAGTGTGGGGGAAAAATGATCTT 4055
DB 1201 TTAGCAAGTCTCAAGTGCAAAATTAACAGTGTGTAGAGTGTGGGGGAAAAATGATCTT 1260
QY 4056 ATTTTTCCTTACAGGAGATACACACTGTGAATTCATTTCACTGAGAGCCCTGACAT 4115
DB 1261 ATTTTTCCTTACAGGAGATACACACTGTGAATTCATTTCACTGAGAGCCCTGACAT 1320
QY 4116 TCTCCCTAAACATAGTGTGTTTCTTAAACAAAGTTTAAAGTGTAAATTAATTAAT 4175
DB 1321 TCTCCCTAAACATAGTGTGTTTCTTAAACAAAGTTTAAAGTGTAAATTAATTAAT 1380
QY 4176 AAAAAAATTTGCTGTGTCTACTTCAAGCTTGTGTTTATGCCCATTTCATATTTGTGTC 4235
DB 1381 AAAAAAATTTGCTGTGTCTACTTCAAGCTTGTGTTTATGCCATTTCATATTTGTGTC 1440
QY 4236 TGTGTGTAATTCATTAACCTTTTGTATACATTTCTGATGTGTAAATTTGTTCTGTAA 4295
DB 1441 TGTGTGTAATTCATTAACCTTTTGTATACATTTCTGATGTGTAAATTTGTTCTGTAA 1500
QY 4296 ATATCTTATTAAGATTCATTTGTAATTAACATTTGTGTGTG 4340
DB 1501 ATATCTTATTAAGATTCATTTGTAATTAACATTTGTGTGTG 1545

RESULT 2
US-09-023-655-322
Sequence 322, Application US/09023655
Patent No. 6607879
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
APPLICANT: Susan G. Stuart
APPLICANT: Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,655
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 322:
SEQUENCE CHARACTERISTICS:
LENGTH: 1290 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TLYMNOR01
CLONE: 140704
US-09-023-655-322

Query Match 13.6%; Score 596; DB 4; Length 1290;
Best Local Similarity 99.8%; Pred. No. 6.6e-133;
Matches 607; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 847 TCCTTGCTGGATTGAAACGTATGAAAGGCGATGCGATGTAAGAGATTAACAGAGA 906
DB 1 TCCTTGCTGGATTGAAACGTATGAAAGGCGATGCGATGTAAGAGATTAACAGAGA 60

QY 907 AAGAAAGCACAAGCTGTTTCAGGCTATGAACTGAAGAGAGAGAAACAGAGCTAT 966
DB 61 AAGAAAGCACAAGCTGTTTCAGGCTATGAACTGAAGAGAGAGAAACAGAGCTAT 120

QY 967 CTGAGAAATTTAACTGAGTGCAGCCGAGCTTTCCGAGACATCCCACTCTGCTC 1026
DB 121 CTGAGAAATTTAACTGAGTGCAGCCGAGCTTTCCGAGACATCCCACTCTGCTC 180

QY 1027 CGAAGCCCTTCATGAGGCGAGAGT-GGAAAGGAGATTAAGAAATCCCATTTTGA 1085
DB 181 CGAAGCCCTTCATGAGGCGAGAGTGGAAAGGAGATTAAGAAATCCCATTTTGA 240

QY 1086 AGTACAGAAAGAAAGACTCTGTTTAAAGAGCTGCTCTGAATTTTCAAAAGTCAAAACA 1145
DB 241 AGTACAGAAAGAAAGACTCTGTTTAAAGAGCTGCTCTGAATTTTCAAAAGTCAAAACA 300

QY 1146 AAAACTCTTAAGACCTCTCTATTAAAGAGAACTCTGCTCTTATCTGAACTGTT 1205
DB 301 AAAACTCTTAAGACCTCTCTATTAAAGAGAACTCTGCTCTTATCTGAACTGTT 360

QY 1206 TGTAAACGTGAATTGAGGAGCCAAAGAAACCCCAAGAAAGCCCGGTCTTCAAGTGAAC 1265
DB 361 TGTAAACGTGAATTGAGGAGCCAAAGAAACCCCAAGAAAGCCCGGTCTTCAAGTGAAC 420

QY 1266 CCACCAAGACTCTCACTCTCCCAAAAGGAGCCAGACCCATCCCAAGAGAAAGCTTC 1325
DB 421 CCACCAAGACTCTCACTCTCCCAAAAGGAGCCAGACCCATCCCAAGAGAAAGCTTC 480

QY 1326 TCAAGTGAATAGAAAGATTGCGGTACCTTTCCACAGAAATGATTTGTGCTTGG 1385
DB 481 TCAAGTGAATAGAAAGATTGCGGTACCTTTCCACAGAAATGATTTGTGCTTGG 540

QY 1386 CACCAAGCTCCCAATACCGTTTACATTAAGGGAATCTCTCCAAAGAGAGAGACT 1445
DB 541 CACCAAGCTCCCAATACCGTTTACATTAAGGGAATCTCTCCAAAGAGAGAGACT 600

QY 1446 GTAGCAAG 1453
DB 601 GTAGCAAG 608

RESULT 3
US-09-702-705-1554
Sequence 1554, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Patrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane

APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1554
LENGTH: 542
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(542)
OTHER INFORMATION: n = A,T,C or G
US-09-702-705-1554

Query Match 11.7%; Score 510.2; DB 4; Length 542;
Best Local Similarity 98.0%; Pred. No. 1.7e-112;
Matches 532; Conservative 4; Mismatches 4; Indels 3; Gaps 2;

QY 3328 AAGGAATTTATTCGAGACATGTAATTTCTTGATGATCTTGCTCTTATTTCT 3387
DB 1 AAGGAATTTATTCGAGACATGTAATTTCTTGATGATCTTGCTCTTATTTCT 60

QY 3388 CTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3447
DB 61 CTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 118

QY 3448 TTAGAGAGTGGGCTCTCTATAAGGAACTGCTGTAACTTCATTCAGCAAGATGTA 3507
DB 119 TTAGAGAGTGGGCTCTCTATAAGGAACTGCTGTAACTTCATTCAGCAAGATGTA 178

QY 3508 GAGGAATTTGACTTAATTCCTAGGAGCTCTCATCTCAACCTTAAGAGAGATTT 3567
DB 179 GAGGAATTTGACTTAATTCCTAGGAGCTCTCATCTCAACCTTAAGAGAGATTT 238

QY 3568 CTGAAAGAACTGGGCGAGATTTCTTGTCTCCATCATTTTAATGAGGAGCTGTCA 3627
DB 239 CTGAAAGAACTGGGCGAGATTTCTTGTCTCCATCATTTTAATGAGGAGCTGTCA 298

QY 3628 GTTTTCTTACTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3687
DB 299 GTTTTCTTACTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358

QY 3688 CAATCAGTCTCTTGAATTTGCTTGAATCCCTCAATTTCTTGAATTTGATGATGAT 3747
DB 359 CAATCAGTCTCTTGAATTTGCTTGAATCCCTCAATTTCTTGAATTTGATGATGAT 418

QY 3748 GT-CGGGTTCTTAATTTTGGGATGATGATGATGATGATGATGATGATGATGATGATGAT 3806
DB 419 GT-CGGGTTCTTAATTTTGGGATGATGATGATGATGATGATGATGATGATGATGATGAT 478

QY 3807 CAGGGAGCTCCCAATTTCTTGAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 3866
DB 479 CAGGGAGCTCCCAATTTCTTGAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 538

QY 3867 TGG 3869
DB 539 TGG 541

RESULT 4
US-09-736-457-1554
Sequence 1554, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom

RESULT 15
US-09-616-289-48
Sequence 48, Application US/09616289
Patent No. 6632923
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
TITLE OF INVENTION: ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/616,289
CURRENT FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/517,849
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26

;
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 2561
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (246)...(1895)
US-09-616-289-48

Query Match 1.5%; Score 64.8; DB 4; Length 2561;
Best Local Similarity 49.4%; Pred. No. 4.6e-05;
Matches 252; Conservative 0; Mismatches 252; Indels 6; Gaps 3;

QY 229 AGCTAAGAGAGCCCGGCGCCCTGCGCTCTCCAGAGCGGAGCCCGCGCTTCCC 288
DB 526 ACCGCAACGGGGCGCGCTCAAGCCCGCGCGGAGCACCCCGCGCGCGCGC 585
QY 289 CGGCCGCTGCGCGCGCAAGAGCGGAGCTTGTACTCCCGCGGAGCGCGCGCGC 348
DB 586 GC 645
QY 349 AGCAGAGAGAGCTGGGCGGCTTCCGTCCTTGCCTTGCCTTGCCTTGCCTTGC 408
DB 646 CGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 705
QY 409 AAGCGCGCATTTGGGGGAGCGCTGCGCGAGCGGAGCGGCTG--CAGCGCGCGCGCA 465
DB 706 CGCGCGCTGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 765
QY 466 AGTATCAGCGGCTGCTGCGCATTCAGACGGGCTCTCTGCGCGCGCGCGCAAGAGCTTA 525
DB 766 AGCGGGC 825
QY 526 CGCGCTGGGCTGGGAGCAAGAGTGGGAGCGCTCCCGCGCTGCACCGCGCTGGAACCGG 585
DB 826 CCCGC 885
QY 586 CGGAGACCCCACTAATCTTGCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 645
DB 886 CGCTGCTGCTGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 944
QY 646 GGAACCTTGGCGGC--CAGCGAGGAGAGATGAGAGAGATGAGAGAGAGCCCTCTCGGAGG 703
DB 945 CCAACCGC 1004
QY 704 TGGTGGCGGCTCGGAGCCTCCAGTCAGGC 733
DB 1005 CCGC 1034

Search completed: March 24, 2005, 11:49:37
Job time : 673 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2005, 05:02:20 ; Search time 18289 Seconds
(without alignments)
11583.258 Million cell updates/sec

Title: US-10-054-935-1

Perfect score: 4372
Sequence: 1 cagtcctcgaccccccacac.....aaaaaaaaaaaaaaaa 4372

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4708233 seqs, 24227607955 residues

Word size : 15

Total number of hits satisfying chosen parameters: 713025

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ste:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2888	66.1	3425	6	C0842718 Sequence
2	2888	66.1	3425	6	AK125654 Homo sapi
3	2643	60.5	2745	9	HSMB807332
4	2511	57.4	194780	9	AC068669 Homo sapi
5	2245	51.3	2296	9	AK055378 Homo sapi
6	2040	46.7	2638	9	AK124185 Homo sapi
7	1260	28.8	1509	6	BD136391 95 human
8	903	20.7	1545	6	AR379622 Sequence
9	564	12.9	655	6	C0724830 Sequence
10	536	12.3	1269	9	BC039449 Homo sapi
11	487	11.1	1290	6	AR379777 Sequence
12	469	10.7	910	6	BD017821 Novel gen
13	469	10.7	910	6	BD097759 Novel gen
14	449	10.3	904	6	BD017820 Novel gen
15	449	10.3	904	6	BD097758 Novel gen
16	436	10.0	600	6	C0706413 Sequence
17	427	9.8	893	6	BD018137 Novel gen
18	427	9.8	893	6	BD098075 Novel gen
19	422	9.7	580	11	BV167952 sqnm6889

20	422	9.7	1092	11	BV178745	BV178745 sqnm10234
21	410	9.4	451	6	CQ713127	CQ713127 Sequence
22	306	7.0	903	6	BD018136	BD018136 Novel gen
23	306	7.0	903	6	BD098074	BD098074 Novel gen
24	281	6.4	400	6	C0663907	C0663907 Sequence
25	240	5.5	240	11	G22752	G22752 human STS W
26	239	5.5	245	6	CQ698720	CQ698720 Sequence
27	219	5.0	2933	10	BC058629	BC058629 Mus muscu
28	214	4.9	440	6	C0687002	C0687002 Sequence
29	152	3.5	542	6	AR273811	AR273811 Sequence
30	152	3.5	542	6	AR277392	AR277392 Sequence
31	152	3.5	542	6	AR407667	AR407667 Sequence
32	152	3.5	542	6	AR441517	AR441517 Sequence
33	152	3.5	542	6	AR544328	AR544328 Sequence
34	152	3.5	542	6	AX368844	AX368844 Sequence
35	134	3.1	211	6	C0524598	C0524598 Sequence
36	127	2.9	196724	10	AL590963	AL590963 Mouse DNA
37	119	2.7	106773	2	AC119462_3	Continuation (4 of
38	119	2.7	183618	2	AC131356	AC131356 Rattus no
39	119	2.7	230913	2	AC119484	AC119484 Rattus no
40	116	2.7	2087	10	BC055715	BC055715 Mus muscu
41	116	2.7	2095	10	BC043039	BC043039 Mus muscu
42	110	2.5	615	6	AX401049	AX401049 Sequence
43	106	2.4	317	6	AX913485	AX913485 Sequence
44	106	2.4	317	6	BD049018	BD049018 Sequence
45	85	1.9	201	11	BV207720	BV207720 sqnm22440
46	60	1.4	60	6	C0552685	C0552685 Sequence
47	55	1.3	208620	2	AC076643	AC076643 Mus muscu
48	51	1.2	590	6	AX400340	AX400340 Sequence
49	51	1.2	615	10	BC043668	BC043668 Mus muscu
50	47	1.1	50	6	AX899390	AX899390 Sequence
51	47	1.1	50	6	BD034923	BD034923 Sequence
52	40	0.9	312459	2	AL928900	AL928900 Danilo rer
53	39	0.9	158220	10	AC140799	AC140799 Mus muscu
54	39	0.9	158220	2	AC102314	AC102314 Mus muscu
55	39	0.9	324084	2	AC133452	AC133452 Mus muscu
56	38	0.9	138	6	AR418098	AR418098 Sequence
57	38	0.9	138	6	AX978792	AX978792 Sequence
58	38	0.9	138	6	BD113651	BD113651 EST and e
59	38	0.9	411	9	BC070483	BC070483 Homo sapi
60	38	0.9	1347	5	BC074531	BC074531 Xenopus t
61	38	0.9	1459	8	ZM066404	U6404 Zee mayus pl
62	38	0.9	1482	10	BC024496	BC024496 Mus muscu
63	38	0.9	1590	9	BC004226	BC004226 Homo sapi
64	38	0.9	1680	9	BC075101	BC075101 Xenopus t
65	38	0.9	1781	5	BC075101	BC075101 Homo sapi
66	38	0.9	1801	5	BC064254	BC064254 Xenopus t
67	38	0.9	2566	9	BC012805	BC012805 Homo sapi
68	38	0.9	4000	9	BC035618	BC035618 Homo sapi
69	38	0.9	102448	10	AL928696	AL928696 Mouse DNA
70	38	0.9	121435	5	BX664613	BX664613 Zebrafish
71	38	0.9	135060	9	AL354659	AL354659 Human DNA
72	38	0.9	142908	2	AL513172	AL513172 Homo sapi
73	38	0.9	152549	2	AL513172	AL513172 Sus scrofa
74	38	0.9	165311	5	AC097061	AC097061 Homo sapi
75	38	0.9	174480	9	AL928975	AL928975 Zebrafish
76	38	0.9	189983	2	AC133524	AC133524 Mus muscu
77	38	0.9	196078	10	AL663091	AL663091 Mouse DNA
78	38	0.9	207691	2	AC119176	AC119176 Mus muscu
79	38	0.9	235259	2	AC118945	AC118945 Rattus no
80	38	0.9	272937	2	AC099162	AC099162 Rattus no
81	38	0.9	280598	2	AC140310	AC140310 Mus muscu
82	37	0.8	100	6	AR421277	AR421277 Sequence
83	37	0.8	100	6	AX981971	AX981971 Sequence
84	37	0.8	100	6	BD116830	BD116830 EST and e
85	37	0.8	243	6	CQ524791	CQ524791 Sequence
86	37	0.8	310	9	HSU88357	HSU88357 Homo sapien
87	37	0.8	446	6	CQ415625	CQ415625 Sequence
88	37	0.8	447	10	BC012238	BC012238 Mus muscu
89	37	0.8	459	6	AR022283	AR022283 Sequence
90	37	0.8	467	6	CQ426545	CQ426545 Sequence
91	37	0.8	651	5	AF425740	AF425740 Danilo rer
92	37	0.8	663	10	BC002062	BC002062 Mus muscu

93 37 0.8 737 10 BC083131
94 37 0.8 753 9 HUMENAE
95 37 0.8 769 10 BC083076
96 37 0.8 772 5 BC082224
97 37 0.8 809 8 BT009076
98 37 0.8 850 5 OSRNATH2
99 37 0.8 863 10 BC047068
100 37 0.8 881 8 AY219049

ALIGNMENTS

RESULT 1
C0842718
LOCUS C0842718 3425 bp DNA linear PAT 02-AUG-2004
DEFINITION Sequence 1365 from Patent EPI440981.
ACCESSION C0842718
VERSION C0842718.1 GI:50894505
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J., Isono,Y., Nagai,K. and Irie,R.
TITLE Full-length human cdna
JOURNAL Patent: EP 1440981-A 1365 28-JUL-2004;
Research Association for Biotechnology (JP)
FEATURES
source 1..3425
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 66.1%; Score 2888; DB 6; Length 3425;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2888; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1453 GGTCGTCATGTCATCAAGTGTGACAGAGAACTTCAGTCTTGCTGCTCTTCTTGA 1512
DB 538 GGTCGTCATGTCATCAAGTGTGACAGAGAACTTCAGTCTTGCTGCTCTTCTTGA 597
QY 1513 GGGACCACTGAGTAGAGCCCTTAAGGGAACCAAACTCTTCAAGCTTTTGGAGAACTGG 1572
DB 598 GGGACCACTGAGTAGAGCCCTTAAGGGAACCAAACTCTTCAAGCTTTTGGAGAACTGG 657
QY 1573 ATGACAGTGTGTTTTCGAACCGGATGCAAACTGAGCTGATGAGAGAGAGAGAGAAA 1632
DB 658 ATGACAGTGTGTTTTCGAACCGGATGCAAACTGAGCTGATGAGAGAGAGAGAGAAA 717
QY 1633 GATGGATATTCAGAGGATCAGGGAACAAGAAATTTTACAGCGACTGAGCTCAGAAATG 1692
DB 718 GATGGATATTCAGAGGATCAGGGAACAAGAAATTTTACAGCGACTGAGCTCAGAAATG 777
QY 1693 ATAAAAAGAAAGAAATTCAGAAATCTGAGCTGAGGTTACTCTATTTTCCCTGAGCAG 1752
DB 778 ATAAAAAGAAAGAAATTCAGAAATCTGAGCTGAGGTTACTCTATTTTCCCTGAGCAG 837
QY 1753 ATGATGTTGAAAGTTTATGATTAACCCCTTCTTGCTGTTGATGCAATTTTGGAGACCAT 1812
DB 838 ATGATGTTGAAAGTTTATGATTAACCCCTTCTTGCTGTTGATGCAATTTTGGAGACCAT 897
QY 1813 TACCAAAATTAATCCACAGAAATTTTGAAGTACCCTGTTGGATGGAGAGCCGATGCA 1872
DB 898 TACCAAAATTAATCCACAGAAATTTTGAAGTACCCTGTTGGATGGAGAGCCGATGCA 957
QY 1873 GATTGAGATCCAGAGAGACAAACACTCAACCGAGCTGTAGAGAAATAGCTGTGCTGGC 1932
DB 958 GATTGAGATCCAGAGAGACAAACACTCAACCGAGCTGTAGAGAAATAGCTGTGCTGGC 1017

QY 1933 AAGAACCTGTCCTTCAGATAGTTGTAGCAATGCCATTCCGAGATGGCAGAGACTGTAT 1992
DB 1018 AAGAACCTGTCCTTCAGATAGTTGTAGCAATGCCATTCCGAGATGGCAGAGACTGTAT 1077
QY 1993 ATGTGACCTTTGTCTCACAATATGTATGACATCTGCTGATATATACCTTTTCAATCTTCT 2052
DB 1078 ATGTGACCTTTGTCTCACAATATGTATGACATCTGCTGATATATACCTTTTCAATCTTCT 1137
QY 2053 GACTTTGTTTCACTTACTGATTTCACAAAAACCTTTTCAATTCGGCTAATTTGTAGTTA 2112
DB 1138 GACTTTGTTTCACTTACTGATTTCACAAAAACCTTTTCAATTCGGCTAATTTGTAGTTA 1197
QY 2113 TGGAGGGTATGAGGATTTCTTTCCCTTTTGGGAAATGGGCTCAAGCTAAAGCTA 2172
DB 1198 TGGAGGGTATGAGGATTTCTTTCCCTTTTGGGAAATGGGCTCAAGCTAAAGCTA 1257
QY 2173 TAGATGAGCAGATTCAGAAAGTTTCAGGGGCTGTTTCTATACATTTTGGCTATGTAAGG 2232
DB 1258 TAGATGAGCAGATTCAGAAAGTTTCAGGGGCTGTTTCTATACATTTTGGCTATGTAAGG 1317
QY 2233 GGTAAAAAGGCTCTTTCATTTAGACATGTGGAAGATGGAAGACAGCCCTTCTTAAAGCT 2292
DB 1318 GGTAAAAAGGCTCTTTCATTTAGACATGTGGAAGATGGAAGACAGCCCTTCTTAAAGCT 1377
QY 2293 GTGCTGCAATGGCACTCTTCTCAACCCCTGGTACACCCCTCTATATAGTGTATGATTT 2352
DB 1378 GTGCTGCAATGGCACTCTTCTCAACCCCTGGTACACCCCTCTATATAGTGTATGATTT 1437
QY 2353 TTAACTCTTAAATTAACCAACCACTCAACATGAGCTTTAGAGCAAGAGAGAAATGAC 2412
DB 1438 TTAACTCTTAAATTAACCAACCACTCAACATGAGCTTTAGAGCAAGAGAGAAATGAC 1497
QY 2413 AAGTGAAGCATTAAGCAAGCCATCTTTCACAGATGTGAAAAAGACATGCGAGAGTTGATG 2472
DB 1498 AAGTGAAGCATTAAGCAAGCCATCTTTCACAGATGTGAAAAAGACATGCGAGAGTTGATG 1557
QY 2473 ATTAAGTGTGAAAAAGATGTTGTTGATTTGAATCTCTGATGATACATGATGAGGA 2532
DB 1558 ATTAAGTGTGAAAAAGATGTTGTTGATTTGAATCTCTGATGATACATGATGAGGA 1617
QY 2533 AGGAGATGTTGGCTGATATTTTTCAGTTAATGATTAACATTTCTTACTGCTCAA 2592
DB 1618 AGGAGATGTTGGCTGATATTTTTCAGTTAATGATTAACATTTCTTACTGCTCAA 1677
QY 2593 AAACCAAAATCTTTGGAAGAAAGTGGGATGTTTGAATTTCAAGCAAGTACAGCTGT 2652
DB 1678 AAACCAAAATCTTTGGAAGAAAGTGGGATGTTTGAATTTCAAGCAAGTACAGCTGT 1737
QY 2653 AAACCAAAAGCACTTAATTTGGGATGGCATGCAAAACCTGATTAATGCTTGTATC 2712
DB 1738 AAACCAAAAGCACTTAATTTGGGATGGCATGCAAAACCTGATTAATGCTTGTATC 1797
QY 2713 ACATCACTTCTCAAGATTTCTCTCATTTGGGCTTCATCTTTTGAAGAACTCTTGTGATG 2772
DB 1798 ACATCACTTCTCAAGATTTCTCTCATTTGGGCTTCATCTTTTGAAGAACTCTTGTGATG 1857
QY 2773 GGGATGAGACTTAAGGAGGTTAGGGGAGAGTGTGAAATAGTGTCTTCTTGGCTGG 2832
DB 1858 GGGATGAGACTTAAGGAGGTTAGGGGAGAGTGTGAAATAGTGTCTTCTTGGCTGG 1917
QY 2833 CAATATGCTACATCTTGAAGCAACAGATGTACTTAATGAGCTTCTTCAATCTTTGTA 2892
DB 1918 CAATATGCTACATCTTGAAGCAACAGATGTACTTAATGAGCTTCTTCAATCTTTGTA 1977
QY 2893 AAATATATTTGTATGTATGATCATCTTGGTCTCTCCCTCCGCTTTGTAAATATCAG 2952
DB 1978 AAATATATTTGTATGTATGATCATCTTGGTCTCTCCCTCCGCTTTGTAAATATCAG 2037
QY 2953 GATAGCACTTCCAGGCACTTTGTGTCTCAGTGTAAATCCCTTAATCTATCTGAAAGGA 3012
DB 2038 GATAGCACTTCCAGGCACTTTGTGTCTCAGTGTAAATCCCTTAATCTATCTGAAAGGA 2097
QY 3013 AAATAGAGCCAAAGCTCTGGTCTCAAAATATATAGAAATTTGCTTTTATGATCTCAGG 3072

```

Db      2098 AATATAGAGCCAAAGCCTCTGCTCAAAATATATAGGAATGACCCTTCTTAACTTCAGG 2157
Qy      3073 ACATATGTGTGAAAACAAAGTAGAGGGTCTATATCTCTGAAGGATGAGGGCTTTATCTCTTA 3132
Db      2158 ACTATGTGTGTGAAAACAAAGTAGAGGGTCTATATCTCTGAAGGATGAGGGCTTTATCTCTTA 2217
Qy      3133 AAGAGATATGTGTGTGAAAACAAAGTAGAGGGTCTATATCTCTGAAGGATGAGGGTGTG 3192
Db      2218 AAGAGATATGTGTGTGAAAACAAAGTAGAGGGTCTATATCTCTGAAGGATGAGGGTGTG 2277
Qy      3193 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3252
Db      2278 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2337
Qy      3253 AAAAGTTCCCTCAGAGGGGCTCCCACTGCTAAAGTTTTTGTGAGATGTGTGTGTGTGTGTGT 3312
Db      2338 AAAAGTTCCCTCAGAGGGGCTCCCACTGCTAAAGTTTTTGTGAGATGTGTGTGTGTGTGTGT 2397
Qy      3313 CTGGATTTGACTTTTAAAGAAATTAATCTGTGCAGACATGATATCTTGTGATGATCTT 3372
Db      2398 CTGGATTTGACTTTTAAAGAAATTAATCTGTGCAGACATGATATCTTGTGATGATCTT 2457
Qy      3373 GCTGCTCTTAATTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3432
Db      2458 GCTGCTCTTAATTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2517
Qy      3433 TGTATCTCCCATCTGTCTTGAAGAGATGAGGCTCTCTATTAAGGAACCTGCTGTAAACCTTCAT 3492
Db      2518 TGTATCTCCCATCTGTCTTGAAGAGATGAGGCTCTCTATTAAGGAACCTGCTGTAAACCTTCAT 2577
Qy      3493 TGTAGCAGAGATGTAGAGAGAAATAGAGACTTAATTTCCACTAGAGGGCTCTCATCTCAGACC 3552
Db      2578 TGTAGCAGAGATGTAGAGAGAAATAGAGACTTAATTTCCACTAGAGGGCTCTCATCTCAGACC 2637
Qy      3553 TTTAAGAGAGATTTCTAGAAAACTGGGCGAGATTTCTTTTGTCTCCATCATTTTAT 3612
Db      2638 TTTAAGAGAGATTTCTAGAAAACTGGGCGAGATTTCTTTTGTCTCCATCATTTTAT 2697
Qy      3613 GTGGCAGGCTGTTCAGTTTCTTACTCTTACTCTATGTGATATTTCTTCGTAACTGTGCCA 3672
Db      2698 GTGGCAGGCTGTTCAGTTTCTTACTCTTACTCTATGTGATATTTCTTCGTAACTGTGCCA 2757
Qy      3673 AAAAGAAAAGAGCCCAATCAGTGTCTCTGACTTGTCTTGTATCCCTCAGTTTCTTC 3732
Db      2758 AAAAGAAAAGAGCCCAATCAGTGTCTCTGACTTGTCTTGTATCCCTCAGTTTCTTC 2817
Qy      3733 TTGATTTCAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3792
Db      2818 TTGATTTCAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2877
Qy      3793 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3852
Db      2878 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2937
Qy      3853 CAGAGGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3912
Db      2938 CAGAGGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2997
Qy      3913 GAGGACCTCTTAATTTTATTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3972
Db      2998 GAGGACCTCTTAATTTTATTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3057
Qy      3973 TGTAGAGAGAGTTGTAGACATGTAGTAAAGTTCCAAAGTGTGAAATTTACAGTGTGTAG 4032
Db      3058 TGTAGAGAGAGTTGTAGACATGTAGTAAAGTTCCAAAGTGTGAAATTTACAGTGTGTAG 3117
Qy      4033 AGTGTGGGGGAAATAGTCTTATTTTCTCTCATGAGGATTAACAACATCTGTGAATTCAA 4092
Db      3118 AGTGTGGGGGAAATAGTCTTATTTTCTCTCATGAGGATTAACAACATCTGTGAATTCAA 3177
Qy      4093 TCTTCAACTGAAGCCCTGCAAGTTCTCTCTTAAACAATAGTTGTGTGTGTGTGTGTGTGTGTGTGT 4152

```

```

Db      3178 TCTTCAACTGAAGCCCTGCAAGTTCTCTTAAACAATAGTTGTGTGTGTGTGTGTGTGTGTGTGT 3237
Qy      4153 GTTTAAGTCTAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4212
Db      3238 GTTTAAGTCTAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3297
Qy      4213 TATGCCATTTCATATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4272
Db      3298 TATGCCATTTCATATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3357
Qy      4273 GTGTAAATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4332
Db      3358 GTGTAAATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3417
Qy      4333 TGGCTGTT 4340
Db      3418 TGGCTGTT 3425

RESULT 2
AKI25654
LOCUS
DEFINITION Homo sapiens cDNA FLJ43666 fis, clone SYN0V4007012.
ACCESSION AKI25654
VERSION AKI25654.1 GI:34531819
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Oca,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Magatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahara,K., Masuno,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3425)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7
Kazuo-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB) ; cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

FEATURES
source
1..3425
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SYNOV4007012"
/tissue_type="synovial membrane tissue from rheumatoid
arthritis"
/clone_id="SYNOV4"
/note="Cloning vector: pME18SFL3"

ORIGIN
Query Match 66.1%; Score 2888; DB 9; Length 3425;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2888; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1453 GGTGTCTGATGTCATCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1512
Db      538 GGTGTCTGATGCCATCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 597

```

QY 1513 GGGACCACTAGTAGAGCCTCTAAAGGACCMAATCCTTCAAGACCTTTGGAGAACTGG 1572
|||
DB 598 GGGACCACTAGTAGAGCCTCTAAAGGACCMAATCCTTCAAGACCTTTGGAGAACTGG 657
|||
QY 1573 ATGACAGTGTGTTTGGAGCGGCATGCAAAACGTGAGCTGGATGAGAGAGAGAAA 1532
|||
DB 658 ATACAGTGTGTTTGGAGCGGCATGCAAAACGTGAGCTGGATGAGAGAGAGAAA 717
|||
QY 1633 GATGGATATTCAAGAGATCAGGGAACAAAGATTTTACAGCGACTGACCTCAGATGT 1692
|||
DB 718 GATGGATATTCAAGAGATCAGGGAACAAAGATTTTACAGCGACTGACCTCAGATGT 777
|||
QY 1693 ATAAAAAGAAAGAAATTCAGGAATCTGAGCCTGAGTTACTCTATTTTCCCTGAGCCAG 1752
|||
DB 778 ATAAAAAGAAAGAAATTCAGGAATCTGAGCCTGAGTTACTCTATTTTCCCTGAGCCAG 837
|||
QY 1753 ATGATGTGAAAGTTGATGATTAACCCCTTCTTGCCGTGTGTAGCATTTGGACGACAT 1812
|||
DB 838 ATGATGTGAAAGTTGATGATTAACCCCTTCTTGCCGTGTGTAGCATTTGGACGACAT 897
|||
QY 1813 TACCAAAATTAATCTCACAAGAAATTTTGAAGTACCTGTTGATGAGCGTAGCCGATGCA 1872
|||
DB 898 TACCAAAATTAATCTCACAAGAAATTTTGAAGTACCTGTTGATGAGCGTAGCCGATGCA 957
|||
QY 1873 GATTGAGATCCAGAGAGACAAACACTCAACGAGCGTGTAGAGAAATAGCTGTGCTGGC 1932
|||
DB 958 GATTGAGATCCAGAGAGACAAACACTCAACGAGCGTGTAGAGAAATAGCTGTGCTGGC 1017
|||
QY 1933 AAGAACCCCTGCTTCAGATAGTTGTAGCATGCGCATTCGCCGAGGTGGCAGAGACCTGTAT 1992
|||
DB 1018 AAGAACCCCTGCTTCAGATAGTTGTAGCATGCGCATTCGCCGAGGTGGCAGAGACCTGTAT 1077
|||
QY 1993 ATGTGACCTTGTCTCCACATATGTTATCACTCGCTGATATATACCTTTGATCTTCTT 2052
|||
DB 1078 ATGTGACCTTGTCTCCACATATGTTATCACTCGCTGATATATACCTTTGATCTTCTT 1137
|||
QY 2053 GACTTGTGTTTCATTAATCTGATTTCAAAAACTCTTTCATTCGCTAATGTGAGTTA 2112
|||
DB 1138 GACTTGTGTTTCATTAATCTGATTTCAAAAACTCTTTCATTCGCTAATGTGAGTTA 1197
|||
QY 2113 TGGAGGGTGATGGGATTTCTTTTCCCTTTTGGGAAATGGGGCTCAAGCTAAACCTA 2172
|||
DB 1198 TGGAGGGTGATGGGATTTCTTTTCCCTTTTGGGAAATGGGGCTCAAGCTAAACCTA 1257
|||
QY 2173 TAGAGTGCGAGATTCAGAGGTTTCAGGGGCTGTGTTCTATATCATTTGCTATGTAAAG 2232
|||
DB 1258 TAGAGTGCGAGATTCAGAGGTTTCAGGGGCTGTGTTCTATATCATTTGCTATGTAAAG 1317
|||
QY 2233 GGTAAAAAGGCTCTCTTCATTTAGACATGTGGAGATGAGCAGCCCTTCTTTAGAGCT 2292
|||
DB 1318 GGTAAAAAGGCTCTCTTCATTTAGACATGTGGAGATGAGCAGCCCTTCTTTAGAGCT 1377
|||
QY 2293 GTGCGTGCGATGGCACTCTTCTCACCCCTGGTACACCCCTTATATGTGGATATAGTATTT 2352
|||
DB 1378 GTGCGTGCGATGGCACTCTTCTCACCCCTGGTACACCCCTTATATGTGGATATAGTATTT 1437
|||
QY 2353 TTAAACCTTAATAATAAACAACAACCTCACCATGAGCTTTAGACCAAGAGAGATGAC 2412
|||
DB 1438 TTAAACCTTAATAATAAACAACAACCTCACCATGAGCTTTAGACCAAGAGAGATGAC 1497
|||
QY 2413 AAGTGAAGCATGAGAGCAAGCCATCTTACAGAGTAGAAAAAGACATCGAGAGTTGTAG 2472
|||
DB 1498 AAGTGAAGCATGAGAGCAAGCCATCTTACAGAGTAGAAAAAGACATCGAGAGTTGTAG 1557
|||
QY 2473 ATTAACGTCTGAAAAAGATAGTTGTTCAATTGAAACTATTCGTGATACAGCATGTGGGA 2532
|||
DB 1558 ATTAACGTCTGAAAAAGATAGTTGTTCAATTGAAACTATTCGTGATACAGCATGTGGGA 1617
|||
QY 2533 AGGAGATTTGGCTGATATTTTTCAGTTAATGTGATAAATTTCTTACTGCTCAA 2592
|||
DB 1618 AGGAGATTTGGCTGATATTTTTCAGTTAATGTGATAAATTTCTTACTGCTCAA 1677
|||

QY 2593 AAACCAAAATCTTTGGAAAAAGAAAGTGGGATGTTAGTTTCAGAAACAATTAACAGTGT 2652
|||
DB 1678 AAACCAAAATCTTTGGAAAAAGAAAGTGGGATGTTAGTTTCAGAAACAATTAACAGTGT 1737
|||
QY 2653 AAACCAAAAGCACTTAGATTTTGGGATGAGCATGCAAAACCTGATATTAATGTCTTGTATC 2712
|||
DB 1738 AAACCAAAAGCACTTAGATTTTGGGATGAGCATGCAAAACCTGATATTAATGTCTTGTATC 1797
|||
QY 2713 ACATCACTTCTCAAGTATTCCTTCAATTGGGCTTCATCTTTTACAGAACTTGTGTGT 2772
|||
DB 1798 ACATCACTTCTCAAGTATTCCTTCAATTGGGCTTCATCTTTTACAGAACTTGTGTGT 1857
|||
QY 2773 GGGATGAGCACTTAGGAGGGGTAGGGGAGAGTGTGAAAAATAGGTCTTCTTGGCTGG 2832
|||
DB 1858 GGGATGAGCACTTAGGAGGGGTAGGGGAGAGTGTGAAAAATAGGTCTTCTTGGCTGG 1917
|||
QY 2833 CAATGTCTACATCTTGAAGAACAAACAGATGTACCTATAGAGCTTCTCATTTCTTGA 2892
|||
DB 1918 CAATGTCTACATCTTGAAGAACAAACAGATGTACCTATAGAGCTTCTCATTTCTTGA 1977
|||
QY 2893 AAAATAATTTGTATGTATACATCTTGTCTCTCCCTCCGTTTGTAAAAATATAG 2952
|||
DB 1978 AAAATAATTTGTATGTATACATCTTGTCTCTCCCTCCGTTTGTAAAAATATAG 2037
|||
QY 2953 GATAGCACTCCAGGCGCATTTGGTCTCAAGTGAAGATCCCTATTTAATCTAGAAAGA 3012
|||
DB 2038 GATAGCACTCCAGGCGCATTTGGTCTCAAGTGAAGATCCCTATTTAATCTAGAAAGA 2097
|||
QY 3013 AAATAGAGCAAGCACTGTGTCTCAAAATATATAGGAATGGCTTCTTATAGCTTCAGG 3072
|||
DB 2098 AAATAGAGCAAGCACTGTGTCTCAAAATATATAGGAATGGCTTCTTATAGCTTCAGG 2157
|||
QY 3073 ACTATTTGTGAAAAACAAGTGGGGTCTAATCTCTAGAAAGTAGGGCTTTTATCTTGA 3132
|||
DB 2158 ACTATTTGTGAAAAACAAGTGGGGTCTAATCTCTAGAAAGTAGGGCTTTTATCTTGA 2217
|||
QY 3133 AAGAGATATATGTCCCAAGATTTATAGCACTTTTAGAGAGAAACCAAGATATAGGGTG 3192
|||
DB 2218 AAGAGATATATGTCCCAAGATTTATAGCACTTTTAGAGAGAAACCAAGATATAGGGTG 2277
|||
QY 3193 TGTGGCTGGCCCATCAAGTGAGACAGAGAGAGAAATGGGATACATTTGTGGAAAGAGAG 3252
|||
DB 2278 TGTGGCTGGCCCATCAAGTGAGACAGAGAGAGAAATGGGATACATTTGTGGAAAGAGAG 2337
|||
QY 3253 AAAAGTTCCTCAGAGGCTTCCCATGCTAAAGTTTGTGAGATGTGATCTGTGCTTC 3312
|||
DB 2338 AAAAGTTCCTCAGAGGCTTCCCATGCTAAAGTTTGTGAGATGTGATCTGTGCTTC 2397
|||
QY 3313 CTGAGATTGACCTTTTAAAGGAATTTCTGGCAGCACATGTATATCTTGGATGATCTT 3372
|||
DB 2398 CTGAGATTGACCTTTTAAAGGAATTTCTGGCAGCACATGTATATCTTGGATGATCTT 2457
|||
QY 3373 GCTGCTCTTATTTCTGCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3432
|||
DB 2458 GCTGCTCTTATTTCTGCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2517
|||
QY 3433 TGTAACTCCATCTGCTTAAAGAGAGTGGGCTCTCTAATAGGGAACCTGCTAAACCTTCAT 3492
|||
DB 2518 TGTAACTCCATCTGCTTAAAGAGAGTGGGCTCTCTAATAGGGAACCTGCTAAACCTTCAT 2577
|||
QY 3493 TGCAGCAAGGATGTAGAGAGAAATAGACCTTAATTCACATAGGGGCTCTCATCTCACACC 3552
|||
DB 2578 TGCAGCAAGGATGTAGAGAGAAATAGACCTTAATTCACATAGGGGCTCTCATCTCACACC 2637
|||
QY 3553 TTAAAGAGAGATTTCTAGAAAAACGTGGGCAATTTTCTTGTCTCCATCAATTTTAAT 3612
|||
DB 2638 TTAAAGAGAGATTTCTAGAAAAACGTGGGCAATTTTCTTGTCTCCATCAATTTTAAT 2697
|||
QY 3613 GTGGCAGGCTGTTCAGTTTTCTTACTCTTACCTATGATATTTCTTCGTAACGTGTCCA 3672
|||
DB 2698 GTGGCAGGCTGTTCAGTTTTCTTACTCTTACCTATGATATTTCTTCGTAACGTGTCCA 2757
|||
QY 3673 AAAAGAAAAAGAACCAATCAAGTGTCTTGACTTGTGTTGTGTGTGTGTGTGTGTGTGTGT 3732
|||


```
Db 2758 AAAAGAAAAAGAACCCAAATCAAGTCTCTGACCTTCTTGTATCCCTCAGTTCTTC 2817
Qy 3733 TTGATTTACAGATGTCGCGGTTCTTAATTTTGGGTATGAGTTAGCAATTAACATG 3792
Db 2818 TTGATTTACAGATGTCGCGGTTCTTAATTTTGGGTATGAGTTAGCAATTAACATG 2877
Qy 3793 TGTATTGACCTTACCCAGGGAGCTCCCAAGTTTCTGACTTGAATAGACTGAGAGATC 3852
Db 2878 TGTATTGACCTTACCCAGGGAGCTCCCAAGTTTCTGACTTGAATAGACTGAGAGATC 2937
Qy 3853 CAGAGGTGCTATCTGGCCAGATTAAAGTATTTCTTCTGTTCTCTCTCTCT 3912
Db 2938 CAGAGGTGCTATCTGGCCAGATTAAAGTATTTCTTCTGTTCTCTCTCTCTCT 2997
Qy 3913 GAGGACCTCTTATTTTATGTCCTCTCTGAGTTTATCTCTCTGATTGACTTGT 3972
Db 2998 GAGGACCTCTTATTTTATGTCCTCTCTGAGTTTATCTCTCTGATTGACTTGT 3057
Qy 3973 TGAGAGAGAGTTGACAGTATAGTACCAAGTTCCAAAGTCCAAATTAACAGTGTAG 4032
Db 3058 TGAGAGAGAGTTGACAGTATAGTATAGCAAGTTCCAAAGTCCAAATTAACAGTGTAG 3117
Qy 4033 AGTGGGGGGGAAATTAATGTTTATTTTCCCTACATGGGATACACACTGTGAATCAA 4092
Db 3118 AGTGGGGGGGAAATTAATGTTTATTTTCCCTACATGGGATACACACTGTGAATCAA 3177
Qy 4093 TCTTCACTGAAGGCGCTGAGTCTCTTAAATATGATGTTGTTTCTTTTAAACAA 4152
Db 3178 TCTTCACTGAAGGCGCTGAGTCTCTTAAATATGATGTTGTTTCTTTTAAACAA 3237
Qy 4153 GTTAAAGCTAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4212
Db 3238 GTTAAAGCTAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3297
Qy 4213 TATGCCATTTCATATGTTGTCGTGTGTATTCATTAATTTGATACATTTCTGAT 4272
Db 3298 TATGCCATTTCATATGTTGTCGTGTGTATTCATTAATTTGATACATTTCTGAT 3357
Qy 4273 GTGTAATAATGCTGTCTGTGTAATATCTTAATAAGTTCATTTGTAATAATCTATG 4332
Db 3358 GTGTAATAATGCTGTCTGTGTAATATCTTAATAAGTTCATTTGTAATAATCTATG 3417
Qy 4333 TGGCTGT 4340
Db 3418 TGGCTGT 3425
```

```
RESULT 3
LSM807332 2745 bp mRNA linear PRI 30-AUG-2003
LOCUS Homo sapiens mRNA; cDNA DKFZp686P24239 (from clone DKFZp686P24239).
ACCESSION BX647188 GI:34366216
VERSION Homo sapiens (human)
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE Homo sapiens Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Eukaryota; Metazoa; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2745)
Wambutt,R., Heubner,D., Neues,H.W., Weill,B., Amid,C., Oeanger,A.,
Fodo,G., Har,M. and Wiemann,S.
The German Human cDNA Consortium
Direct Submission
Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuberberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
consortium of the German Genome Project.
This clone (DKFZp686P24239) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
```

Information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cdna/>.

FEATURES

source

1..2745

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp686P24239"

/issue_type="human small intestine"

/clone_id="686 (synonym: hlcc3). Vector pSport1_Sfi; host

DH10B; sites SfiI + SfiIb"

/dev_stage="adult"

2707..2712

polya_signal

polya_site

2726

ORIGIN

Query Match 60.5%; Score 2643; DB 9; Length 2745;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2743; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy 1615 ATGAGAGAGAGAGAGAGAGATGATATTCAGAGATCAGGAGCAAGAAATTTTACAGC 1674
Db 1 ATGAGAGAGAGAGAGAGAGATGATATTCAGAGATCAGGAGCAAGAAATTTTACAGC 60
Qy 1675 GACTGACGCTCAGATGATATTAAGAGAGAGAGATTCAGAGATCAGAGCTGAGTTACT 1734
Db 61 GACTGACGCTCAGATGATATTAAGAGAGAGAGATTCAGAGATCAGAGCTGAGTTACT 120
Qy 1735 CATTTTCCCTGAGCCAGATGATGTTGAAAGTTGATGATTAACCCCTTCTGCTGTTG 1794
Db 121 CATTTTCCCTGAGCCAGATGATGTTGAAAGTTGATGATTAACCCCTTCTGCTGTTG 180
Qy 1795 TAGCATTTGAGACGACATTAACCAAAATTAATCCAGAGATTTGAGTACCCGTTGG 1854
Db 181 TAGCATTTGAGACGACATTAACCAAAATTAATCCAGAGATTTGAGTACCCGTTGG 240
Qy 1855 ATGAGCGTACCGATGAGATGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914
Db 241 ATGAGCGTACCGATGAGATGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Qy 1915 GGAATAGCTGTGTGTGCAAGAAACCTGTCTTCAGATGTTGATGATGCCATTCCGAG 1974
Db 301 GGAATAGCTGTGTGTGCAAGAAACCTGTCTTCAGATGTTGATGATGCCATTCCGAG 360
Qy 1975 AGTGGCAGAGACCTGTATGATGACCTTGTCCCATATGATATGATCGCGATAT 2034
Db 361 AGTGGCAGAGACCTGTATGATGACCTTGTCCCATATGATATGATCGCGATAT 420
Qy 2035 ACCCTTCATACCTCTTGAATTTGTTTCTTATTAATCTGATTTGACAAAATCTTTTAT 2094
Db 421 ACCCTTCATACCTCTTGAATTTGTTTCTTATTAATCTGATTTGACAAAATCTTTTAT 480
Qy 2095 TCGGCTAATGATGATGATGAGAGGATTTGGATTTCTTTTCCCTTTTGGGAAATGG 2154
Db 481 TCGGCTAATGATGATGATGAGAGGATTTGGATTTCTTTTCCCTTTTGGGAAATGG 540
Qy 2155 GCTCTCAGCTAAAGCTATGATGAGATTCAGAGATTCAGAGATTCAGAGGCTCTTTCTATC 2214
Db 541 GCTCTCAGCTAAAGCTATGATGAGATTCAGAGATTCAGAGATTCAGAGGCTCTTTCTATC 600
Qy 2215 ATTGCTATGATTAAGAGGATTAAGAGGCTCTCTTCAATTAAGACATGAGAGATGAAGA 2274
Db 601 ATTGCTATGATTAAGAGGATTAAGAGGCTCTCTTCAATTAAGACATGAGAGATGAAGA 660
Qy 2275 GCCCTTCCTTGAAGCTGTGCTGACATGACATCTTCTCAACCTGTATACCTCTCTTA 2334
Db 661 GCCCTTCCTTGAAGCTGTGCTGACATGACATCTTCTCAACCTGTATACCTCTCTTA 720
Qy 2335 TAGGGGTATAGATATTTTAAACCTTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2394
Db 721 TAGGGGTATAGATATTTTAAACCTTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy 2395 GACCAAGAGAGATGACAGTGAAGAGATGAAGCAAGCATTTTCAAGAGTGAAGAAG 2454
```

Db 781 GACGAGAAAGGAAATGACAAAGCGATGAAGCAAGCCATCTTCAAGATGAGAAAG 840
Qy 2455 ACATCGAAGAGTTGGTAGATTAAGTCTGAAAAAGATAGTTTCATTGGAACATATCTG 2514
Db 841 ACATCGAAGAGTTGGTAGATTAAGTCTGAAAAAGATAGTTTCATTGGAACATATCTG 900
Qy 2515 TGATTAAGTCATGTGGGAAGGAGATGTTTGGCTGTGATTTATTTTTCAGTTAAAGATAC 2574
Db 901 TGATTAAGTCATGTGGGAAGGAGATGTTTGGCTGTGATTTATTTTTCAGTTAAAGATAC 960
Qy 2575 AATTTCTTACTGCTCAAAAACCAAAATCTTTGAAAAAGAAAGTGGGGATGTTAGTTTC 2634
Db 961 AATTTCTTACTGCTCAAAAACCAAAATCTTTGAAAAAGAAAGTGGGGATGTTAGTTTC 1020
Qy 2635 AGAACAAGTTTACAGCTGTAAACAAAACCACTTAGTATTGGATGCGATGCCAAACCTG 2694
Db 1021 AGAACAAGTTTACAGCTGTAAACAAAACCACTTAGTATTGGATGCGATGCCAAACCTG 1080
Qy 2695 TATTAATGTCTGTATACATCACTTCTCAAGTATTCCTTCATTTGGGCTTCATCCTTTT 2754
Db 1081 TATTAATGTCTGTATACATCACTTCTCAAGTATTCCTTCATTTGGGCTTCATCCTTTT 1140
Qy 2755 AGCAGAACTCTGTGTGGGATAGAGCTTAGGAGGGATAGGGGGAGATGAGAAATA 2814
Db 1141 AGCAGAACTCTGTGTGGGATAGAGCTTAGGAGGGATAGGGGGAGATGAGAAATA 1200
Qy 2815 GGTGCTTCTTTGGCTGGCAAAATGTCTACATCTTGAACAAACAGATGTACTAATAGC 2874
Db 1201 GGTGCTTCTTTGGCTGGCAAAATGTCTACATCTTGAACAAACAGATGTACTAATAGC 1260
Qy 2875 TTCTCAATTCATCTTTGTAATAATTTGTAATGTATGTAACATCTTGGCTCTCCCTCC 2934
Db 1261 TTCTCAATTCATCTTTGTAATAATTTGTAATGTATGTAACATCTTGGCTCTCCCTCC 1320
Qy 2935 GTTTGTTAAATATACAGATAGACCTCCAGGCACTTGTGCTCAGTGAAGATCCT 2994
Db 1321 GTTTGTTAAATATACAGATAGACCTCCAGGCACTTGTGCTCAGTGAAGATCCT 1380
Qy 2995 ATTAACATCTGAAGAAATAGAGCCAGCAAGCTCTGTCTCAATATATAGAAATTC 3054
Db 1381 ATTAACATCTGAAGAAATAGAGCCAGCAAGCTCTGTCTCAATATATAGAAATTC 1440
Qy 3055 CTTTCTTATAGTCTTCAGAGATATTTGTGTAAAAACAGTAGGGGCTTAATCTCTAGAAG 3114
Db 1441 CTTTCTTATAGTCTTCAGAGATATTTGTGTAAAAACAGTAGGGGCTTAATCTCTAGAAG 1500
Qy 3115 TAGGGCTTTATCCTTAAGAGATATGTCCAGATTAATTAAGCACTTTTAGAGGAA 3174
Db 1501 TAGGGCTTTATCCTTAAGAGATATGTCCAGATTAATTAAGCACTTTTAGAGGAA 1560
Qy 3175 GCCAAGTATGTAAGGTGTGTGGCTGCCATCAGTGAAGACAGAAAGAAATGGGATA 3234
Db 1561 GCCAAGTATGTAAGGTGTGTGGCTGCCATCAGTGAAGACAGAAAGAAATGGGATA 1620
Qy 3235 CCAATTGGGAGAGAAAGAAAGTTCTCTCAGGGGCTCCCATCTGCTAAAGTTTGTGA 3294
Db 1621 CCAATTGGGAGAGAAAGAAAGTTCTCTCAGGGGCTCCCATCTGCTAAAGTTTGTGA 1680
Qy 3295 GATGTGATGTGCTCTCTGGAATTTGACTTTTAAAGAAATTAATTCGGACAGCATGTA 3354
Db 1681 GATGTGATGTGCTCTCTGGAATTTGACTTTTAAAGAAATTAATTCGGACAGCATGTA 1740
Qy 3355 GATATCTGGATATCTTGTGCTCTTATTTCTCTTTGTGTGTGTGTGTGTGTGTGTGTG 3414
Db 1741 GATATCTGGATATCTTGTGCTCTTATTTCTCTTTGTGTGTGTGTGTGTGTGTGTGTG 1800
Qy 3415 TGCTATATGGTTTTCATTGTAACTCATCTGCTTAGAGAGTGGGCTCTCTATAGGGA 3474
Db 1801 TGCTATATGGTTTTCATTGTAACTCATCTGCTTAGAGAGTGGGCTCTCTATAGGGA 1860
Qy 3475 ACCTGCTGTAACTTCAATTCAGCAAGAGATGTAGAGAAATAGAACTTAATTCACATAG 3534

Db 1861 ACCTGCTGTAACTTCAATTCAGCAAGAGATGTAGAGAAATAGAACTTAATTCACATAG 1920
Qy 3535 GGGCTCTCATCTCAACCTTTAAGAGAGATTTCTAGAAAAAATCTGGGCCAGATTTCTTT 3594
Db 1921 GGGCTCTCATCTCAACCTTTAAGAGAGATTTCTAGAAAAAATCTGGGCCAGATTTCTTT 1980
Qy 3595 GTTCTCATCATTTTAATGTGGCAGGCTGTGCTTTCTTACTCTTAACCTATGTATAT 3654
Db 1981 GTTCTCATCATTTTAATGTGGCAGGCTGTGCTTTCTTACTCTTAACCTATGTATAT 2040
Qy 3655 TTCTTGTAACGTGTCCAAAAAGAAAAAGAACCAATCAGTGTCTGTGACTTTGTTCTT 3714
Db 2041 TTCTTGTAACGTGTCCAAAAAGAAAAAGAACCAATCAGTGTCTGTGACTTTGTTCTT 2100
Qy 3715 TGATCCCTCAGTTCTTCTTGATTGAGCATGTGTGGGTTCTTAATTTGGGTATGAT 3774
Db 2101 TGATCCCTCAGTTCTTCTTGATTGAGCATGTGTGGGTTCTTAATTTGGGTATGAT 2160
Qy 3775 TAGCAATTTTAACATTTGTTTGTGTGCTTACCCAGAGGACCTCCAGTTTCTGACTGA 3834
Db 2161 TAGCAATTTTAACATTTGTTTGTGTGCTTACCCAGAGGACCTCCAGTTTCTGACTGA 2220
Qy 3835 AGTAGACTGAGAGATTCACAGAGTCTATCTGSCACAGTTTAAGTAGATTCATTTCC 3894
Db 2221 AGTAGACTGAGAGATTCACAGAGTCTATCTGSCACAGTTTAAGTAGATTCATTTCC 2280
Qy 3895 TTGGTTCTCCCTCTCCCTGAGACCTTTATTTATTTTTCCTCTTCTAGTTAATCT 3954
Db 2281 TTGGTTCTCCCTCTCCCTGAGACCTTTATTTATTTTTCCTCTTCTAGTTAATCT 2340
Qy 3955 CTTTATTTTGAATTTGTTGAGAGAGAGTTGAGCAGTATAGCAAAATGCCAAGTGC 4014
Db 2341 CTTTATTTTGAATTTGTTGAGAGAGAGTTGAGCAGTATAGCAAAATGCCAAGTGC 2400
Qy 4015 AAATTAACAGTGTGTAGAGTGTGGGGGAAATTAAGTCTTATTTTCCCTACATGGAT 4074
Db 2401 AAATTAACAGTGTGTAGAGTGTGGGGGAAATTAAGTCTTATTTTCCCTACATGGAT 2460
Qy 4075 ACAACACTGTGAATTAATCTTCACTGAAGGCTGTGACAGTCTCTTAACAATAGTTGT 4134
Db 2461 ACAACACTGTGAATTAATCTTCACTGAAGGCTGTGACAGTCTCTCTTAACAATAGTTGT 2520
Qy 4135 TTGTTTTCTTTAACAAAGTTAAGCTAGTGTTAATTAATAAATAATTTGCTGTGCTG 4194
Db 2521 TTGTTTTCTTTAACAAAGTTAAGCTAGTGTTAATTAATAAATAATTTGCTGTGCTG 2580
Qy 4195 TCTACTTACGCTTTGTTTATAGCCATTTGATATTTGTGTGCTGTGTATCATTAAT 4254
Db 2581 TCTACTTACGCTTTGTTTATAGCCATTTGATATTTGTGTGCTGTGTATCATTAAT 2640
Qy 4255 TTGTATACATTTCTGATGTGTAATAATGTGTTGCTTGTAAATATCTTATTAAGATTCA 4314
Db 2641 TTGTATACATTTCTGATGTGTAATAATGTGTTGCTTGTAAATATCTTATTAAGATTCA 2700
Qy 4315 ATTGTAAATTAATCTATTTGTGGCTGTAAAAAATTTTAAAAAATTTTAAAAA 4359
Db 2701 ATTGTAAATTAATCTATTTGTGGCTGTAAAAAATTTTAAAAAATTTTAAAAA 2745

RESULT 4
AC068669 194780 bp DNA linear PRI 17-JAN-2003
LOCUS Homo sapiens chromosome 17, clone RP11-749I16, complete sequence.
DEFINITION AC068669
AC068669.21 GI:27777612
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 194780)
Birren, B., Nussbaum, C. and Lander, E.
Homo sapiens chromosome 17, clone RP11-749I16

JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 194780)
Birren,B., Linton,J., Barna,N., Bastien,V., Bada,F.,
Anderson,S., Boukhalter,B., Brown,A., Burkett,G.,
Boguslavsky,L., Castelle,A., Choepel,Y., Colangelo,M., Collins,S.,
Campoliano,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kam,L., Karakas,A.,
Klein,R., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,T.,
Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.,J.,
Young,G., Zainoun,U., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Direct Submission
Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 194780)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camazata,J., Chang,J., Chazaro,B., Choepel,Y., Collimore,A.,
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gind,S., Graham,L., Grand-Pierre,N., Hafetz,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karakas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J.,
Mathews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tessfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Direct Submission
Submitted (14-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 194780)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camazata,J., Chang,J., Chazaro,B., Choepel,Y., Collimore,A.,
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gind,S., Graham,L., Grand-Pierre,N., Hafetz,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karakas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J.,
Mathews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tessfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT
Submitted (17-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 17, 2003 this sequence version replaced g1:27734042.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L10045
Center clone name: 749_I_16

FEATURES
Source
Location/Qualifiers
1..194780
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-749116"
/clone_11b="RPC1-11 Human Male BAC"
2..285
/rpt_family="L1MC3"
302..324
/rpt_family="TTG)n"
529..637
/rpt_family="FLAM_A"
548..553
/note="<30 qual SINGL region"
complement(711..1007)
/rpt_family="AluSx"
complement(1038..1179)
/rpt_family="MER115"
2169..2473
/rpt_family="AluSp"
3642..3946
/rpt_family="AluY"
complement(4181..4391)
/rpt_family="AluIo"
4863..4901
/rpt_family="GGGA)n"
5391..5716
/rpt_family="AluSx"
6700..6739
/rpt_family="CA)n"
complement(7332..7491)
/rpt_family="MIR3"
7684..7805
/rpt_family="CGGG)n"
7700..7856
/note="single clone coverage"
7788..7802
/note="<30 qual single clone coverage"
7825..7845
/note="<30 qual single clone coverage"
7931..7951
/rpt_family="GC_rich"
7980..8009
/rpt_family="GC_rich"
8388..8442
/rpt_family="GC_rich"
9670..9780
/rpt_family="L2"
10423..10557
/rpt_family="CT-rich"
11371..11675
/rpt_family="AluSx"
11724..11913
/rpt_family="MIR"
12708..13030
/rpt_family="AluIb"
13109..13189
/rpt_family="MIR3"
14044..14165
/rpt_family="CT-rich"

repeat_region 14184..14282
/rpt_family="CT-rich"
repeat_region 15860..15976
/rpt_family="CT-rich"
repeat_region 15894..16062
/rpt_family="C-rich"
repeat_region 17522..17592
/rpt_family="C-rich"
repeat_region 18316..18420
/rpt_family="MIR"
repeat_region 21106..21188
/rpt_family="MIR"
repeat_region 22950..22950
/rpt_family="AluSx"
repeat_region 23005..23172
/rpt_family="MIR"
repeat_region 23173..23481
/rpt_family="AluSx"
repeat_region 23431..23431
/note="<30 qual SINGL region"
repeat_region 23482..23525
/rpt_family="MIR"
repeat_region 23629..23697
/rpt_family="MIRSA"
repeat_region 24000..24000
/rpt_family="AluSc"
repeat_region 24001..24113
/rpt_family="MIRSA"
repeat_region 24215..24483
/rpt_family="AluSx"
repeat_region 24529..24534
/note="<30 qual SINGL region"
repeat_region 24541..24551
/note="<30 qual SINGL region"
repeat_region 24561..24569
/note="<30 qual SINGL region"
repeat_region 24586..24632
/rpt_family="AluSx"
repeat_region 24644..24770
/rpt_family="MIR"
repeat_region 24821..25072
/rpt_family="AluSx"

Query Match 57.4%; Score 2511; DB 9; Length 194780;
Best Local Similarity 100.0%; Pred.No.0;
Matches 2511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1832 GAATTTGAGCTACCCGTGTTGATGAGCGTAGCCGATGAGATTGGAGATCCAGAA 1891
DB 79250 GAATTTGAGCTACCCGTGTTGATGAGCGTAGCCGATGAGATTGGAGATCCAGAA 79309
QY 1892 GCAACACCTCAGCCGAGCGTAGGAATAGCTGTGCGCAAGAACCTGTCTTCAGAT 1951
DB 79310 GCAACACCTCAGCCGAGCGTAGGAATAGCTGTGCGCAAGAACCTGTCTTCAGAT 79369
QY 1952 AGTTGAGCATGCGCATCCCGAGAGTGGCAGAGACCTGTATATGTGACCTTTGTCTCAC 2011
DB 79370 AGTTGAGCATGCGCATCCCGAGAGTGGCAGAGACCTGTATATGTGACCTTTGTCTCAC 79429
QY 2012 ATATGTTATCAGTCGCGATATACCTTTCAATCTTCCTTGAATTTGTTTCAATTC 2071
DB 79430 ATATGTTATCAGTCGCGATATACCTTTCAATCTTCCTTGAATTTGTTTCAATTC 79489
QY 2072 TGAATTCACAAAACCTTTCAATCTTCCTTGAATTTGTTTCAATTC 2131
DB 79490 TGAATTCACAAAACCTTTCAATCTTCCTTGAATTTGTTTCAATTC 79549
QY 2132 CTTTTCCTTTTGGGAAATGGGCTCTCAAGCTAAAGCTATAGAGATGGCAGATTGAGAA 2191
DB 79550 CTTTTCCTTTTGGGAAATGGGCTCTCAAGCTAAAGCTATAGAGATGGCAGATTGAGAA 79609
QY 2192 GTTTCAGGGGTCTGTTCTATACATTTGCTATGTTAAAGGGGTAAAGGCTCTCTCA 2251

DB 79610 GTTTCAGGGGTCTGTTCTATACATTTGCTATGTTAAAGGGGTAAAGGCTCTCTCA 79669
QY 2252 TTAGACATGTGGAATGAAAGCAGCCCTTCTTTAGAGCTGTGCTGCATGGCAGCTTT 2311
DB 79670 TTAGACATGTGGAATGAAAGCAGCCCTTCTTTAGAGCTGTGCTGCATGGCAGCTTT 79729
QY 2312 CTCACCTGGTACACCTCTTATAGTGGGTATAGATTTTAAACCTTAAATTAACA 2371
DB 79730 CTCACCTGGTACACCTCTTATAGTGGGTATAGATTTTAAACCTTAAATTAACA 79789
QY 2372 AACAACTCAGATGAGCTTTAGACAGAAAGAGATGACAAAGATGAGTAAAGCAA 2431
DB 79790 AACAACTCAGATGAGCTTTAGACAGAAAGAGATGACAAAGATGAGTAAAGCAA 79849
QY 2432 GCCATCTTACAGAGTAAAGAACATCGAGAGTTGTATGATTAATCTGTGTAAGATA 2491
DB 79850 GCCATCTTACAGAGTAAAGAACATCGAGAGTTGTATGATTAATCTGTGTAAGATA 79909
QY 2492 GTTGTTCATTTGAACTATCTGTGATACATGATGGAAGGATGTTGGCTGTGAT 2551
DB 79910 GTTGTTCATTTGAACTATCTGTGATACATGATGGAAGGATGTTGGCTGTGAT 79969
QY 2552 TATTTTTCAGTTAATGATTAACAATTTCTTTACTGCTCAAAAACCAAAATCTTTGAAA 2611
DB 79970 TATTTTTCAGTTAATGATTAACAATTTCTTTACTGCTCAAAAACCAAAATCTTTGAAA 80029
QY 2612 AGAAAGTGGGAGATGTTAGTTTACAGAACAGTTACAGCTGTAAACAAAGACCTTAGAT 2671
DB 80030 AGAAAGTGGGAGATGTTAGTTTACAGAACAGTTACAGCTGTAAACAAAGACCTTAGAT 80089
QY 2672 TTGGAGATGGAGTGCAGAAACCTGTATTAATGTCCTGTATACATCACTCTCAAGATTT 2731
DB 80090 TTGGAGATGGAGTGCAGAAACCTGTATTAATGTCCTGTATACATCACTCTCAAGATTT 80149
QY 2732 CTTTCAATTTGGCTTACCTTTTACAGAACTCTTGTGTGGATAGAGACTTAGGAG 2791
DB 80150 CTTTCAATTTGGCTTACCTTTTACAGAACTCTTGTGTGGATAGAGACTTAGGAG 80209
QY 2792 GGTAGGGGAGAGTGGAAATAGTGTCTCTTGTGCTGGCAATGTCTACATCTTGAA 2851
DB 80210 GGTAGGGGAGAGTGGAAATAGTGTCTCTTGTGCTGGCAATGTCTACATCTTGAA 80269
QY 2852 ACAACAGATATCTTAATAGCTTCTCAATCTTTGTAATTAATTTGATATGTA 2911
DB 80270 ACAACAGATATCTTAATAGCTTCTCAATCTTTGTAATTAATTTGATATGTA 80329
QY 2912 CCATCTTGTCTCTCCCTCCGCTTTGTTAAATATACAGATAGCACTCCAGGCTCAC 2971
DB 80330 CCATCTTGTCTCTCCCTCCGCTTTGTTAAATATACAGATAGCACTCCAGGCTCAC 80389
QY 2972 TTTGTCTCAGTAAAGATCCTATTATCTATCTGAAGAAATAGAGCCAAAGCTCT 3031
DB 80390 TTTGTCTCAGTAAAGATCCTATTATCTATCTGAAGAAATAGAGCCAAAGCTCT 80449
QY 3032 GGTCTCAAAATATAGAAATGTGCTTTTATCTTACCTTACAGACTATGTGTGAAACAAAG 3091
DB 80450 GGTCTCAAAATATAGAAATGTGCTTTTATCTTACCTTACAGACTATGTGTGAAACAAAG 80509
QY 3092 TAGGGGTCTATCTCTAGAAAGTAGGGGCTTTATCTTAAAGAAATATGTCCCGAGA 3151
DB 80510 TAGGGGTCTATCTCTTAAAGAGTAGGGGCTTTATCTTAAAGAAATATGTCCCGAGA 80569
QY 3152 TTATTAACATTTTAAAGAGAACCAAGATATGAGGTGTGTGCTGAGCCATCAGTG 3211
DB 80570 TTATTAACATTTTAAAGAGAACCAAGATATGAGGTGTGTGCTGAGCCATCAGTG 80629
QY 3212 GAGCAGAAAGAGATGGATACATTTGTGGAAAGAAAGTTCTCAGGGGCT 3271
DB 80630 GAGCAGAAAGAGATGGATACATTTGTGGAAAGAAAGTTCTCAGGGGCT 80689
QY 3272 CCGACGTCTAAAGTTTGTGAGATGTGATCTGTGCTCTCGATTTGACTTTTAAAG 3331
DB 80690 CCGACGTCTAAAGTTTGTGAGATGTGATCTGTGCTCTCGATTTGACTTTTAAAG 80749

/clone="FEBR2001571"
/issue_type="brain"
/clone_id="FEBR2"
/dev_stage="fetus"
/note="Cloning vector: pME18SFL3"

Query Match 51.3%; Score 2245; DB 9; Length 2296;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 2295; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN

2024 TCGCTAATAATACCTTTTCATCTCCTTACCTTGTGTTTCATTTCTATCTGATTTCAAAA 2083
1 TCGCTAATAATACCTTTTCATCTCCTTACCTTGTGTTTCATTTCTATCTGATTTCAAAA 60
2084 AACTCTTTCATTTGCGCTAATTTGTGAGTTATGAGGGTATGAGGATTTCTTTTCCCTTTT 2143
61 AACTCTTTCATTTGCGCTAATTTGTGAGTTATGAGGGTATGAGGATTTCTTTTCCCTTTT 120
2144 TTGGGAAATGGGCTCTCAAGCTAAAGCTATAGATAGGACATTCAGAAAGTTTCAGGGGTC 2203
121 TTGGGAAATGGGCTCTCAAGCTAAAGCTATAGATAGGACATTCAGAAAGTTTCAGGGGTC 180
2204 TGTTCCTATACATTTGCTGCTATGTTAAAGGGGTAAAGGGCTCTCTTCATTAAGACATGG 2263
181 TGTTCCTATACATTTGCTGCTATGTTAAAGGGGTAAAGGGCTCTCTTCATTAAGACATGG 240
2264 AAGATGAAGAGCGCCCTCTCTTAAAGCTGTGCTGCTGATGCACTCTTTCACCTTGTA 2323
241 AAGATGAAGAGCGCCCTCTCTTAAAGCTGTGCTGCTGATGCACTCTTTCACCTTGTA 300
2324 CACCCCTCTTATAGTGGTATAGTATTTTAAACCTTAAATAAACAAACAACTTACC 2383
301 CACCCCTCTTATAGTGGTATAGTATTTTAAACCTTAAATAAACAAACAACTTACC 360
2384 ATGAGCTTTTGAACAAGAGATGACAAAGTGAAGCGATGAAGCAAGCATCTTACA 2443
361 ATGAGCTTTTGAACAAGAGATGACAAAGTGAAGCGATGAAGCAAGCATCTTACA 420
2444 GAGTAGAAAAGACATCGAGAGTTGTGATAACTGTCTGAAAAGATAGTTGTCATTTG 2503
421 GAGTAGAAAAGACATCGAGAGTTGTGATAACTGTCTGAAAAGATAGTTGTCATTTG 480
2504 AACTATTTCTGTGATATAGTCAATGTGGGAAGGATGTTGGCTGTATATTTTTCAAT 2563
481 AACTATTTCTGTGATATAGTCAATGTGGGAAGGATGTTGGCTGTATATTTTTCAAT 540
2564 TAAATGATAACAATTTCTTACTGCTCAAAAACCAAAATCTTTTGGAAAAGAAAGTGGGA 2623
541 TAAATGATAACAATTTCTTACTGCTCAAAAACCAAAATCTTTTGGAAAAGAAAGTGGGA 600
2624 TGGTAGTTTCAAGAACAGTTTACAGCTGTAAACAAAACCACTTAATGATTTGGGATGGCAT 2683
601 TGGTAGTTTCAAGAACAGTTTACAGCTGTAAACAAAACCACTTAATGATTTGGGATGGCAT 660
2684 GCCAAAACCTGTATTAATGTCTTGTATCAATCACTTTCAGATTTCTTCAATTTGGGC 2743
661 GCCAAAACCTGTATTAATGTCTTGTATCAATCACTTTCAGATTTCTTCAATTTGGGC 720
2744 TTCACTCTTTTGAAGAACTTGTGTGGTGGATAGACCTTACGAGGGGTAAAGGGGAGA 2803
721 TTCACTCTTTTGAAGAACTTGTGTGGTGGATAGACCTTACGAGGGGTAAAGGGGAGA 780
2804 GTGTGAAATAGAGTCTTCTTGGCTGCAATATGTCTACATCTTTGAAAACAAACAGATGT 2863
781 GTGTGAAATAGAGTCTTCTTGGCTGCAATATGTCTACATCTTTGAAAACAAACAGATGT 840
2864 ACCTAATGAGCTTCTCAATTCATTTGTAAATATTTGTATGTATGATACCATCTTGTGTC 2923
841 ACCTAATGAGCTTCTCAATTCATTTGTAAATATTTGTATGTATGATACCATCTTGTGTC 900
2924 TCTCCCTCCCGTTTCTTAAATATATGATACCATCTCCAGGCCACTTTGTGCTCAAT 2983
|||||

901 TCTCCCTCCCGTTTCTTAAATATATGAGATAGACCTCCAGGCCACTTTGTGCTCAAT 960
2984 GTTAAAGTCCCTTATTAATCTTGAAGAAATATAGGCCAAAGCTCTGTCTCAATAT 3043
961 GTTAAAGTCCCTTATTAATCTTGAAGAAATATAGGCCAAAGCTCTGTCTCAATAT 1020
3044 ATGAAATGGCTTTTCTTATGCTTCAAGACTATGTGTGAAGAAAGATAGGGGTCTAAT 3103
1021 ATGAAATGGCTTTTCTTATGCTTCAAGACTATGTGTGAAGAAAGATAGGGGTCTAAT 1080
3104 CTCTGAAGAGTAAAGGGCTTTTATCTTAAAGAAATATGTCCCAATTTATTTAGCACTT 3163
1081 CTCTGAAGAGTAAAGGGCTTTTATCTTAAAGAAATATGTCCCAATTTATTTAGCACTT 1140
3164 TTGAGAGAAAGCCAAAGTATGAGGTGTGTGTGAGGTGTGTGTGTGTGTGTGTGTGTGT 3223
1141 TTGAGAGAAAGCCAAAGTATGAGGTGTGTGTGAGGTGTGTGTGTGTGTGTGTGTGTGT 1200
3224 AGAATGGGATACCATTTGTGGGAAAGAAAGAAAGTCTCAAGGGGCTCCCACTGTCTAA 3283
1201 AGAATGGGATACCATTTGTGGGAAAGAAAGAAAGTCTCAAGGGGCTCCCACTGTCTAA 1260
3284 GTTTTGTGAGATGTTGATCTGTGCTCTCTGATTTGACTTTTAAAGAAATTTCTGG 3343
1261 GTTTTGTGAGATGTTGATCTGTGCTCTCTGATTTGACTTTTAAAGAAATTTCTGG 1320
3344 CAGCAGATGATTTCTTTGATGATGATCTGTCTCTTATTTCTCTTTTGTGTGTGTGT 3403
1321 CAGCAGATGATTTCTTTGATGATGATCTGTCTCTTATTTCTCTTTTGTGTGTGTGT 1380
3404 GT 3463
1381 GT 1440
3464 TCTATAAGGAACTCTCTGTAACCTTCAATTCAGCAAGATGTAGAGAAATATAGACTT 3523
1441 TCTATAAGGAACTCTCTGTAACCTTCAATTCAGCAAGATGTAGAGAAATATAGACTT 1500
3524 AATTCCAATAGGGGCTCTCATCTCACACCTTAAAGGAGGATTTCTAGAAAACTGGGCC 3583
1501 AATTCCAATAGGGGCTCTCATCTCACACCTTAAAGGAGGATTTCTAGAAAACTGGGCC 1560
3584 AGATTTTCTTGTCTCCATCATTTTAATGTGCAAGCTGTTCAGTTTCTTCTCTCTTAC 3643
1561 AGATTTTCTTGTCTCCATCATTTTAATGTGCAAGCTGTTCAGTTTCTTCTCTCTTAC 1620
3644 CTATGATATTTCTTCTGTAACCTGTCAAAAAAGAAAAAGAACCAATCAGTGTCTTGG 3703
1621 CTATGATATTTCTTCTGTAACCTGTCAAAAAAGAAAAAGAACCAATCAGTGTCTTGG 1680
3704 ACTTTGTTCTTGATCCCTCAAGTTCTTCTTGAATTTCAAGATGTGTGGGTTCTTAATTT 3763
1681 ACTTTGTTCTTGATCCCTCAAGTTCTTCTTGAATTTCAAGATGTGTGGGTTCTTAATTT 1740
3764 TGGGTATAGTTAGCAATTTTAAACCAATGTGTTTGGCTTACCAAGGGGACCTCCCAAT 3823
1741 TGGGTATAGTTAGCAATTTTAAACCAATGTGTTTGGCTTACCAAGGGGACCTCCCAAT 1800
3824 TTCTGACTTGAATGACTGAGAAAGATCCACAGAGTGTATCTGACAGATTTAAAGTAG 3883
1801 TTCTGACTTGAATGACTGAGAAAGATCCACAGAGTGTATCTGACAGATTTAAAGTAG 1860
3884 ATTCTATTTCTTGTGTTCTCTCTCTCTCTGAGAGCTCTTATTTTATTTGTCTCTCTCT 3943
1861 ATTCTATTTCTTGTGTTCTCTCTCTCTCTCTGAGAGCTCTTATTTTATTTGTCTCTCTCT 1920
3944 AGGTTAATTTCTCTTGAATTTGACTTTGTGAGAAAGAGTGTGACATGATTTAGTAA 4003
1921 AGGTTAATTTCTCTTGAATTTGACTTTGTGAGAAAGAGTGTGACATGATTTAGTAA 1980
4004 GTTCCAAGTCAAAATTAACAGTGTGTTAGAGTGTGGGGGAGAAATAGTCTTATTTTCTC 4063
1981 GTTCCAAGTCAAAATTAACAGTGTGTTAGAGTGTGGGGGAGAAATAGTCTTATTTTCTC 2040
|||||

QY	2232	GGGTAAGAAAGGCTCTCTTCATTTAACAATGAGGAAAGATGAAGACACCCCTTCCTTTAAGC	2231
Dp	1199	GGGTAAGAAAGGCTCTCTTCATTTAACAATGAGGAAAGATGAAGACACCCCTTCCTTTAAGC	1258
QY	2292	TGTGCTGCATGAGCACTCTTCTCAACCCCTGTACACCCCTCTTAATAGTGGTATAGATT	2351
Dp	1259	TGTGCTGCATGAGCACTCTTCTCAACCCCTGTACACCCCTCTTAATAGTGGTATAGATT	1318
QY	2352	TTTAAACCTCTAAATTTAAACCAACACTCAACATGAGCTTTTAGGACCAAGAGAGATGA	2411
Dp	1319	TTTAAACCTCTAAATTTAAACCAACACTCAACATGAGCTTTTAGGACCAAGAGAGATGA	1378
QY	2412	CAAGTGAAGCATATAGACAGACCACTCTCAACAGATGAGAAAGACATCGAGAGTTAGTA	2471
Dp	1379	CAAGTGAAGCATATAGACAGACCACTCTCAACAGATGAGAAAGACATCGAGAGTTAGTA	1438
QY	2472	GATATAGTCTGAAAGAAGATAGTTGTCACTTTGAAACTATTCGTGATACATCATGTGG	2531
Dp	1439	GATATAGTCTGAAAGAAGATAGTTGTCACTTTGAAACTATTCGTGATACATCATGTGG	1498
QY	2532	AAGGATGTTGGCTGCTGATATTTTCACTTAATGATTAACAATTCCTTACTAGCTCA	2591
Dp	1499	AAGGATGTTGGCTGCTGATATTTTCACTTAATGATTAACAATTCCTTACTAGCTCA	1558
QY	2592	AAAAACAACATCTTGTGAAAAAGAAAGTGGGATGTTAGTTTCAGAACATGTTACAGCTG	2651
Dp	1559	AAAAACAACATCTTGTGAAAAAGAAAGTGGGATGTTAGTTTCAGAACATGTTACAGCTG	1618
QY	2652	TAAACAAAGACATTAGTATTTTGGGATGCGCATGCCCCATGATTAATATGCTCTGTAT	2711
Dp	1619	TAAACAAAGACATTAGTATTTTGGGATGCGCATGCCCCATGATTAATATGCTCTGTAT	1678
QY	2712	CACATCACTTCTCAAGTATTCCTTCATCTGGGCTTCATCCTTTAGACAACTCTTGAGG	2771
Dp	1679	CACATCACTTCTCAAGTATTCCTTCATCTGGGCTTCATCCTTTAGACAACTCTTGAGG	1738
QY	2772	TGGATATGAGACTTAGGGAGGATAGGGGAGAGTGTGAAATAGAGTCTCTTGAGCTG	2831
Dp	1739	TGGATATGAGACTTAGGGAGGATAGGGGAGAGTGTGAAATAGAGTCTCTTGAGCTG	1798
QY	2832	GCAATATGCTACATCTTGGAAACAAACAGATGACCTATGAGCTTCTCATTCACTTGT	2891
Dp	1799	GCAATATGCTACATCTTGGAAACAAACAGATGACCTATGAGCTTCTCATTCACTTGT	1858
QY	2892	AAAAATATTTGTATGTGTACCATCTTGCTCCTCCCTCCCGTTTGTATTAATAATCA	2951
Dp	1859	AAAAATATTTGTATGTGTACCATCTTGCTCCTCCCTCCCGTTTGTATTAATAATCA	1918
QY	2952	GGATATGCACTCCACAGGCCACTTTGTGTCTCACTGTATAGATCCCTATTAACATCTGAAAG	3011
Dp	1919	GGATATGCACTCCACAGGCCACTTTGTGTCTCACTGTATAGATCCCTATTAACATCTGAAAG	1978
QY	3012	AAAAATAGAGCAAGACCTCTGATCTCAAAATATATGAAATGCGCTTCTTATAGCTTCAG	3071
Dp	1979	AAAAATAGAGCAAGACCTCTGATCTCAAAATATATGAAATGCGCTTCTTATAGCTTCAG	2038
QY	3072	GACTATTTGTGTGAAAACAAGTATGAGGGGTCTTAATCTCTTAGAAGTATGAGGGCTTTATCCTT	3131
Dp	2039	GACTATTTGTGTGAAAACAAGTATGAGGGGTCTTAATCTCTTAGAAGTATGAGGGCTTTATCCTT	2098
QY	3132	AAAGAGAAATATGTCCCGCAGATTATTAAGACCTTTTAGAGAGAAAGCCAGATATGATAGGGT	3191
Dp	2099	AAAGAGAAATATGTCCCGCAGATTATTAAGACCTTTTAGAGAGAAAGCCAGATATGATAGGGT	2158
QY	3192	GTTGTGGCTGGCCCATCACTGTGAGACGAAAGAGAGATGGATATCCATTTGTGGAGAGAA	3251
Dp	2159	GTTGTGGCTGGCCCATCACTGTGAGACGAAAGAGAGATGGATATCCATTTGTGGAGAGAA	2218
QY	3252	GAAAAAGTTCTCAAGGGGCTCCCACTGTGAAAGTTTCTTGTAGATGTTATCTGTGCTT	3311
Dp	2219	GAAAAAGTTCTCAAGGGGCTCCCACTGTGAAAGTTTCTTGTAGATGTTATCTGTGCTT	2278

[illegible]

LOCUS	BD136391	1509 bp	DNA	linear	PAT 18-SEP-2002
DEFINITION	95 human secretory proteins.				
ACCESSION	BD136391				
VERSION	BD136391.1 GI:2321336				
KEYWORDS	JP 2002506627-A/78.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindaes; Homo.				
AUTHORS	Ruben,S.M., Ni,J., Rosen,C.A., Yu,G.L., Young,P.E., Fen,P., Soppe,D.W., Wei,Y.F., Endress,G.A., Duan,R.D., Kyaw,H., Ebner,R., Lafleur,D.W., Olsen,H.S., Shi,Y. and Moore,P.A.				
TITLE	95 human secretory proteins				
JOURNAL	Patent: JP 2002506627-A 78 05-MAR-2002;				
COMMENT	HUMAN GENOME SCIENCES INC				
	OS	Homo sapiens (human)			
	PN	JP 2002506627-A/78			
	PD	05-MAR-2002			
	PF	18-MAR-1999 JP 2000536733			
	PR	19-MAR-1998 US 60/078566,19-MAR-1998 US 60/078576 PR			
		19-MAR-1998 US 60/078573,19-MAR-1998 US 60/078574 PR			
		19-MAR-1998 US 60/078579,19-MAR-1998 US 60/078578 PR			
		19-MAR-1998 US 60/078581,19-MAR-1998 US 60/078577 PR			
		19-MAR-1998 US 60/078583,01-APR-1998 US 60/080314 PR			
		01-APR-1998 US 60/080312,01-APR-1998 US 60/080313 PI			
		M RUBEN,DIAN NI,CRAIG A ROSEN,GUO			
		LIANG YU,PAUL E YOUNG,			
		PING FENG,			
	PI	DANIEL R SOPPET,YING FEI WEI,GREGORY A ENDRESS,ROXANNE D DUAN,			
	PI	HLA KYAW,			
	PI	REINHARD EBNER,DAVID W LAFLEUR,HENRIK S OLSEN,YANGSU SHI,PAUL			
	PI	A MOORE			
	PC	C12N15/09,A61K38/00,A61K48/00,A61P43/00,C07K14/47,C07K16/18,			
	PC	C12N1/15,			
	PC	C12N1/19,C12N1/21,C12N5/10,C12P21/02,C12Q1/68,G01N33/53,G01N33/ PC			
		566,			
	PC	C12N15/00,A61K37/02,C12N5/00			
	CC	95 human secretory proteins			
	FH	Key Location/Qualifiers			
	FT	source 1..1509			
	FT	Location/Qualifiers /organism='Homo sapiens (human)'			
FEATURES					
source	1..1509				
	/organism='Homo sapiens'				

ORIGIN /mol_type="genomic DNA"
/db_xref="taxon:9606"

Query Match 28.8%; Score 1260; DB 6; Length 1509;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
2859 GATGCTATATAGCTTCTCCATTCACCTTGTAAATAATATTGTATGTACCATCTT 2918
9 GATGCTATATAGCTTCTCCATTCACCTTGTAAATAATATTGTATGTACCATCTT 68
2919 GGTCTCTCCCTCCCGTTTGTAAATATCAGATAGACCTCCAGCCACTTTGTC 2978
69 GGTCTCTCCCTCCCGTTTGTAAATATCAGATAGACCTCCAGCCACTTTGTC 128
2979 TCAGTATAGATCCCTATTAATCTGAAAGAAAAATAGAGCAAGCTCTGTCTCA 3038
129 TCAGTATAGATCCCTATTAATCTGAAAGAAAAATAGAGCAAGCTCTGTCTCA 188
3039 AATAATAGGAATTCCTTTCTTTAGCTTCAGAGACTATGTGAAAAACAAGTAGGGGT 3098
189 AATAATAGGAATTCCTTTCTTTAGCTTCAGAGACTATGTGAAAAACAAGTAGGGGT 248
3099 CTAAATCTCTAGAGGATAGGGGCTTTTATCTTAAAGAAATATGTCCCAATTATAG 3158
249 CTAAATCTCTAGAGGATAGGGGCTTTTATCTTAAAGAAATATGTCCCAATTATAG 307
3159 CACTTTTAAGAGAAAGCCAAAGTATGTA-GGGTGTGTGGCTGGCCCATATGTAGACAC 3217
308 CACTTTTAAGAGAAAGCCAAAGTATGTA-GGGTGTGTGGCTGGCCCATATGTAGACAC 367
3218 GAAGAGAGATGGGATACCATTTGGGAGAGAGAAAGAAAGTTCTCAGGGGCTCCCACT 3277
368 GAAGAGAGATGGGATACCATTTGGGAGAGAGAAAGAAAGTTCTCAGGGGCTCCCACT 427
3278 GCTAAAGTTTGTGAGATGTGATCTGTCTTCCCTGATTTGACTTTTAAAGAAATTA 3337
428 GCTAAAGTTTGTGAGATGTGATCTGTCTTCCCTGATTTGACTTTTAAAGAAATTA 487
3338 TTCTGGACAGACATGATATTTCTTGAGATCTTGTCTCTTATTTCTCTTTTGTGT 3397
488 TTCTGGACAGACATGATATTTCTTGAGATCTTGTCTCTTATTTCTCTTTTGTGT 547
3398 GGT 3457
548 GGT 607
3458 GGGCTCTCTAATAGGAACCTGCTGTAACTTCAATTCAGACAGATGTAGAGAAATA 3517
608 GGGCTCTCTAATAGGAACCTGCTGTAACTTCAATTCAGACAGATGTAGAGAAATA 667
3518 GGAATTAATTCACATAGGGGCTCTCATCTCACAACCTTAAAGAGAGATTTCTAGAAAAC 3577
668 GGAATTAATTCACATAGGGGCTCTCATCTCACAACCTTAAAGAGAGATTTCTAGAAAAC 727
3578 TGGGCGAGATTTTCTTGTCTCAGATCTTAAATGTGAGAGGCTGTAGCTTTCTTAC 3637
728 TGGGCGAGATTTTCTTGTCTCAGATCTTAAATGTGAGAGGCTGTAGCTTTCTTAC 787
3638 TCTTACCTAATGATATTTTCTTGTATCAGTGTCCAAAAAAGAAAAAGAACCAATCAGTGT 3697
788 TCTTACCTAATGATATTTTCTTGTATCAGTGTCCAAAAAAGAAAAAGAACCAATCAGTGT 847
3698 CTCTTACCTTGTCTTGTATCAGTGTCTTCTTGTATTCAGATGTGTGGGTTCC 3757
848 CTCTTACCTTGTCTTGTATCAGTGTCTTCTTGTATTCAGATGTGTGGGTTCC 907
3758 TATTTTGGGTATGAGTATGAGTAAATTAACATTTGTTGGTGGCCCAACCAAGGAGATC 3817
908 TATTTTGGGTATGAGTATGAGTAAATTAACATTTGTTGGTGGCCCAACCAAGGAGATC 967
3818 CCCAGTTTCTGACTGATGAGTGAAGTGAAGAAATCCAGAGTGTCTATCTGGCCAGATTT 3877

Db 968 CCCAGTTTCTGACTGATGAGTGAAGTGAAGAAATCCAGAGTGTCTATCTGGCCAGATTT 1027
Qy 3878 AAGTATGATCTAATTTCTTGTGTCTCTCCCTCCCTAGAGACCTCTAATTTATTTATGTCCTCC 3937
Db 1028 AAGTATGATCTAATTTCTTGTGTCTCTCCCTCCCTAGAGACCTCTAATTTATTTATGTCCTCC 1087
Qy 3938 TCTTCTAGGTTAATTTCTTGT 3997
Db 1088 TCTTCTAGGTTAATTTCTTGT 1147
Qy 3998 AGCAAGTTTCAAGTGCAGAAATTAACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4057
Db 1148 AGCAAGTTTCAAGTGCAGAAATTAACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1207
Qy 4058 TTTTCCCTACATGGGATPACACACCTGGAATTCATCTTCAACGTGAAGCCCTGACGTTTC 4117
Db 1208 TTTTCCCTACATGGGATPACACACCTGGAATTCATCTTCAACGTGAAGCCCTGACGTTTC 1267
Qy 4118 TCTTAAACATAGT 4177
Db 1268 TCTTAAACATAGT 1327
Qy 4178 AAAAAATGCTGT 4237
Db 1328 AAAAAATGCTGT 1387
Qy 4238 TGTGTATTTCAATCTTTGT 4297
Db 1388 TGTGTATTTCAATCTTTGT 1447
Qy 4298 ATCTTAAAGATTTCAATCTTTGT 4357
Db 1448 ATCTTAAAGATTTCAATCTTTGT 1507
Qy 4358 AA 4359
Db 1508 AA 1509

RESULT 8
AR379622 1545 bp DNA linear PAT 18-DEC-2003
LOCUS AR379622
DEFINITION Sequence 167 ffrom patent US 6607879.
ACCESSION AR379622
VERSION AR379622.1 GI:40087256
KEYWORDS
SOURCE
ORGANISM
Unclasseified.
REFERENCE 1 (bases 1 to 1545)
AUTHORS Cocks,B.G., Stuart,S.G. and Seilhamer,J.J.
TITLE Compositions for the detection of blood cell and immunological
response gene expression
JOURNAL Patent: US 6607879-A 167 19-AUG-2003;
FEATURES
source location/Qualifiers
1..1545
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 20.7%; Score 903; DB 6; Length 1545;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
3426 TTTCATTTGTAATCTCATCTGTAGAGAGTGGGCTCTCTATAGGGAACCTGCTGTAA 3485
Db 631 TTTCATTTGTAATCTCATCTGTAGAGAGTGGGCTCTCTATAGGGAACCTGCTGTAA 690
Qy 3486 ACTTCATTTGAGAGATGTAGAGAAATTAAGACTTAATTTCACTAGAGGGCTTCATC 3545
Db 691 ACTTCATTTGAGAGATGTAGAGAAATTAAGACTTAATTTCACTAGAGGGCTTCATC 750

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullany, S.J., Bosak, S.A., Mowen, P.J.,
Mokeman, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Vallalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywicki, M.T., Skalska, U., Smailus, D.E.,
Schmechel, A., Schein, J.B., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palkovics, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
Series: IRAP Plate: 77 Row: n Column: 1.

FEATURES

source
1. 1269
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5311571"
/tissue_type="Brain, hypothalamus"
/clone_id="N1H MGC_96"
/lab_host="DH10B"
/note="Vector: pBluescript"

ORIGIN

Query Match 12.3%; Score 536; DB 9; Length 1269;
Best Local Similarity 99.8%; Pred. No. 2,7e-306;
Matches 586; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 867 ATGGAAGCGCGATGCGCTGGTAAAGAGATACGAGAAAGGCAACAGCTGTTT 926
DB 1 ATGGAAGCGCGATGCGCTGGTAAAGAGATACGAGAAAGGCAACAGCTGTTT 60
QY 927 CAGGCGATGAATCTGAAGAGAGAGAGAAACAGAGATCTGAGAAATTTAACTGAG 986
DB 61 CAGGCGATGAATCTGAAGAGAGAGAGAAACAGAGATCTGAGAAATTTAACTGAG 120
QY 987 TGCAGCGGAGCTTTCGAGACATCCAGACTCTGCTCCAGAGCCCTTCTCATGTGG 1046
DB 121 TGCAGCGGAGCTTTCGAGACATCCAGACTCTGCTCCAGAGCCCTTCTCATGTGG 180
QY 1047 CGGAGTGAAGAGACATTAAGAAATCCCATTTTGAAGTACAGAAAGAGATCTCT 1106
DB 181 CGGAGTGAAGAGACATTAAGAAATCCCATTTTGAAGTACAGAAAGAGATCTCT 240

QY 1107 GTTAAAGTGCCTCTGTAATTTTCAAAAGTCAAAACAAAACCTCTAAGCACTCTCT 1166
DB 241 GTTAAAGTGCCTCTGTAATTTTCAAAAGTCAAAACAAAACCTCTAAGCACTCTCT 300
QY 1167 ATTAAGAGAACCTCTGTTCTTATCTGAAGCTTTTGTAACTGTAACTGAAGAGC 1226
DB 301 ATTAAGAGAACCTCTGTTCTTATCTGAAGCTTTTGTAACTGTAACTGAAGAGC 360
QY 1227 CAAGAAACCCAGAAAAGCCCGGTCTTCAAGTGAACACCCCAAGACTCTCCACTCC 1286
DB 361 CAAGAAACCCAGAAAAGCCCGGTCTTCAAGTGAACACCCCAAGACTCTCCACTCC 420
QY 1287 CAAGAGGACCCAGACCCATCCCAAGAGAAAACCTTCTCAAGTGAATGAAGATTG 1346
DB 421 CAAGAGGACCCAGACCCATCCCAAGAGAAAACCTTCTCAAGTGAATGAAGATTG 480
QY 1347 CCGTACCTTTCCACACAGAAATGATTTGTGTCTTGGACACAGCTCTCCCATCAC 1406
DB 481 CCGTACCTTTCCACACAGAAATGATTTGTGTCTTGGACACAGCTCTCCCATCAC 540
QY 1407 TTACCATTAAGGGAATCTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1453
DB 541 TTACCATTAAGGGAATCTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 587

RESULT 11
AR379777
LOCUS AR379777 1290 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 322 from patent US 6607879.
ACCESSION AR379777
VERSION AR379777.1 GI:40087411
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1290)
Cocks, B.G., Stuart, S.G. and Selthamer, J.J.
Compositions for the detection of blood cell and immunological
response gene expression
Patent: US 6607879-A 322 19-AUG-2003;—

FEATURES
source
1. 1290
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 11.1%; Score 487; DB 6; Length 1290;
Best Local Similarity 99.8%; Pred. No. 3.8e-277;
Matches 607; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 847 TCCCTGCTCGGATTTGAACGTATGGAAGGCGGATGCACTGGTAAAGATTAACGAGA 906
DB 1 TCCCTGCTCGGATTTGAACGTATGGAAGGCGGATGCACTGGTAAAGATTAACGAGA 60
QY 907 AAGAAAGGCAAGCTGTTTCAGGGCTATGAATCTGAAGAGAGAGAGAGAGAGAGAT 966
DB 61 AAGAAAGGCAAGCTGTTTCAGGGCTATGAATCTGAAGAGAGAGAGAGAGAGAGAT 120
QY 967 CTGAGAAATTTAACTGAGATGCGAGCGGAGCTTTCGAGACATCCAGACTCTGCTC 1026
DB 121 CTGAGAAATTTAACTGAGATGCGAGCGGAGCTTTCGAGACATCCAGACTCTGCTC 180
QY 1027 CCAAGCCCTTCTCATGTGGGCGAGT GGAAGAGACATTAAGAAATCCCATTTTGA 1085
DB 181 CCAAGCCCTTCTCATGTGGGCGAGT GGAAGAGACATTAAGAAATCCCATTTTGA 240
QY 1086 AGTACGAAGAAAGAGACTCTGTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1145
DB 241 AGTACGAAGAAAGAGACTCTGTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 1146 AAAAATCTTAAG 1205
DB 301 AAAAATCTTAAG 360

```
QY 1206 TGTAAACGTGAATTGAGAGCCAGAAACCCGAAAGAGCCCGCTTTCAGTGAACACC 1265
DB 361 TGTAAACGTGAATTGAGAGCCAGAAACCCGAAAGAGCCCGCTTTCAGTGAACACC 420
QY 1266 CCACCAAGACTCTCCGACTCCCGAAAGGAGCCAGAGCCATCCCAAGAGAGAGGCTTC 1325
DB 421 CCACCAAGACTCTCCGACTCCCGAAAGGAGCCAGAGCCATCCCAAGAGAGAGGCTTC 480
QY 1326 TCAAGTAGATAGAAAGATTTCGCGTACCTTTCACACAGAAATGTAATTTGTGCTGTGG 1385
DB 481 TCAAGTAGATAGAAAGATTTCGCGTACCTTTCACACAGAAATGTAATTTGTGCTGTGG 540
QY 1386 CACCAAGCTTCCCGATCACCCTTACCAATTCGGGAATCCTCTCCAAAGAGAGAGACT 1445
DB 541 CACCAAGCTTCCCGATCACCCTTACCAATTCGGGAATCCTCTCCAAAGAGAGAGACT 600
QY 1446 GTAGCAAG 1453
DB 601 GTAGCAAG 608
```

```
RESULT 12
LOCUS BD017821/c 910 bp DNA linear PAT 27-AUG-2002
DEFINITION Novel gene and novel gene fragment cloned in human neuroblastoma.
ACCESSION BD017821
VERSION BD017821.1 GI:22558997
KEYWORDS JP 2001245671-A/59.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 910)
REFERENCE
AUTHORS Nakagawa, A.
TITLE Novel gene and novel gene fragment cloned in human neuroblastoma
JOURNAL Patent: JP 2001245671-A 59 11-SEP-2001;
CHIBA PREF. HISAMITSU PHARMACEUTICAL CO INC
COMMENT OS Homo sapiens (human)
PN JP 2001245671-A/59
PD 11-SEP-2001
PE 07-MAR-2000 JP 2000159195
PI AKIRA NAKAGAWARA
PC C12N15/09, C12Q1/68, G01N33/53, G01N33/566// (C12Q1/68, C12R1.91),
CC Novel gene and novel gene fragment cloned in human CC
FH Key source Location/Qualifiers
FT source 1..910
FEATURES
source Location/Qualifiers
1..910
/mol_type="genomic DNA"
/db_xref="taxon:9606"
```

```
ORIGIN
Query Match 10.7%; Score 469; DB 6; Length 910;
Best Local Similarity 100.0%; Pred. No. 1.8e-266;
Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3867 TGGCCAGATTAAAGTAGATTCTATTTCTGTTCTCCCTCCCTGAGAGACCTTTATT 3926
DB 520 TGGCCAGATTAAAGTAGATTCTATTTCTGTTCTCCCTCCCTGAGAGACCTTTATT 461
QY 3927 TTAATGCTCCCTCTTCTAGAGTTAATTCCTTTGATTTGACTTTGTTGAGAAGAGGTTG 3986
DB 460 TTAATGCTCCCTCTTCTAGAGTTAATTCCTTTGATTTGACTTTGTTGAGAAGAGGTTG 401
QY 3987 GACAGTAGATTAGCAAAAGTCCAAAGTCCAAATTAACAGTGTGTAGAGTGGGGGGGAAA 4046
DB 400 GACAGTAGATTAGCAAAAGTCCAAAGTCCAAATTAACAGTGTGTAGAGTGGGGGGGAAA 341
```

```
QY 4047 ATTAGCTTAATTTTCCCTACATGGGATACAAACCTGTGAATTCATCTTCACTGAAG 4106
DB 340 ATTAGCTTAATTTTCCCTACATGGGATACAAACCTGTGAATTCATCTTCACTGAAG 281
QY 4107 CCTGAGTTCCTTCAAAACATAGTTGTTGTTTTCTTTAAACAAGTTTAAGCTAGTG 4166
DB 280 CCTGAGTTCCTTCAAAACATAGTTGTTGTTTTCTTTAAACAAGTTTAAGCTAGTG 221
QY 4167 TAATAAATTAATAAATAATGCTTGTCTGTCTACTTCAGCTTTGTTTATGCCATTTCAT 4226
DB 220 TAATAAATTAATAAATAATGCTTGTCTGTCTACTTCAGCTTTGTTTATGCCATTTCAT 161
QY 4227 ATTGTTGCTGCTGCTTGAATTAATCACTTTGATACCACTTTCGAGTGAATAATGGTT 4286
DB 160 ATTGTTGCTGCTGCTTGAATTAATCACTTTGATACCACTTTCGAGTGAATAATGGTT 101
QY 4287 GTCTTGAATAATCTTATTAAGATTCAATTGTAATAAATTAATTAATTTGTGG 4335
DB 100 GTCTTGAATAATCTTATTAAGATTCAATTGTAATAAATTAATTAATTTGTGG 52
```

```
RESULT 13
LOCUS BD097759/c 910 bp DNA linear PAT 27-AUG-2002
DEFINITION Novel genes cloned in humanneuroblastoma and fragments thereof.
ACCESSION BD097759
VERSION BD097759.1 GI:22643333
KEYWORDS WO 0166719-A/59.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 910)
REFERENCE
AUTHORS Nakagawa, A.
TITLE Novel genes cloned in humanneuroblastoma and fragments thereof
JOURNAL Patent: WO 0166719-A 59 13-SEP-2001;
CHIBA PREF. HISAMITSU PHARMACEUTICAL CO INC, AKIRA NAKAGAWARA
COMMENT OS Homo sapiens (human)
PN WO 0166719-A/59
PD 13-SEP-2001
PE 02-MAR-2001 WO 2001JP001629
PR 07-MAR-2000 JP 00P 159195
PI AKIRA NAKAGAWARA
PC C12N15/11, C12Q1/68, G01N33/53, G01N33/566
CC Novel genes cloned in humanneuroblastoma and fragments thereof
FH Key source Location/Qualifiers
FT source 1..910
FEATURES
source Location/Qualifiers
1..910
/mol_type="genomic DNA"
/db_xref="taxon:9606"
```

```
ORIGIN
Query Match 10.7%; Score 469; DB 6; Length 910;
Best Local Similarity 100.0%; Pred. No. 1.8e-266;
Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3867 TGGCCAGATTAAAGTAGATTCTATTTCTGTTCTCCCTCCCTGAGAGACCTTTATT 3926
DB 520 TGGCCAGATTAAAGTAGATTCTATTTCTGTTCTCCCTCCCTGAGAGACCTTTATT 461
QY 3927 TTAATGCTCCCTCTTCTAGAGTTAATTCCTTTGATTTGACTTTGTTGAGAAGAGGTTG 3986
DB 460 TTAATGCTCCCTCTTCTAGAGTTAATTCCTTTGATTTGACTTTGTTGAGAAGAGGTTG 401
QY 3987 GACAGTAGATTAGCAAAAGTCCAAAGTCCAAATTAACAGTGTGTAGAGTGGGGGGGAAA 4046
DB 400 GACAGTAGATTAGCAAAAGTCCAAAGTCCAAATTAACAGTGTGTAGAGTGGGGGGGAAA 341
QY 4047 ATTAGCTTAATTTTCCCTACATGGGATACAAACCTGTGAATTCATCTTCACTGAAG 4106
```

Df		340	AATGAGCTTATTATTTCCCATACATGGGAACAACACTGAAATTCATCCTCAACTGAAG	281
Oy		4107	CCCCGCAGTGTTCTCTAANAACATAAGTAGTTTTGTTCCTTAACAAGAAGTTAAGTAAGT	4166
Df		280	CCCgcagcttccctaAAAacatAgtagtggtttggttttcTTTaAcAAaggTTaaGtaagt	221
Oy		4167	TAAATAAATAAAAAAAATTCGTGTGCTGTCTACTTCAAGCTTGTTATATGCCAATTTCAT	4226
Df		220	TAAATAAATAAAAAAAATTCGTGTGCTGTCTACTTCAAGCTTGTTATATGCCAATTTCAT	161
Oy		4227	ATTGTGTCTGTGTGTATATCATTAACATTTTGATACCATTTCGTATGTGTAAAAATGGTT	4286
Df		160	ATTGTGTCTGTGTGTATATCATTAACATTTTGATACCATTTCGTATGTGTAAAAATGGTT	101
Oy		4287	GTCCTTGTAAATATCTTATTAAGAGTTCATTGTATTAATTAACATTATTTGGG	4335
Df		100	GTCCTTGTAAATATCTTATTAAGAGTTCATTGTATTAATTAACATTATTTGGG	52
RESULT_14				
Bd017820				
LOCUS		BD017820	904 bp DNA linear PAT 27-AUG-2002	
DEFINITION		Novel gene and novel gene fragment cloned in human neuroblastoma.		
ACCESSION		BD017820 .1 GI:22558996		
VERSION		JF 2001245671-A/58.		
KEYWORDS		Homo sapiens (human)		
SOURCE		Homo sapiens		
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. Nakagawara,A. I (bases 1 to 904) Novel gene and novel gene fragment cloned in human neuroblastoma Patent : JP 2001245671-A 58 11-SEP-2001; CHIBA PREP,HISAMITSU PHARMACEUTICAL CO INC		
REFERENCE		OS Homo sapiens (human)		
AUTHORS		PN JP 2001245671-A/58		
TITLE		PD 11-SEP-2001		
JOURNAL		PF 07-MAR-2000 JP 2000159195		
COMMENT		PI AKIRA NAKAGAWARA		
		PC C12N15/09,C12Q1/68,G01N33/53,G01N33/566//(C12Q1/68,C12R1:91), PC C12N15/00		
		CC Novel gene and novel gene fragment cloned in human CC neuroblastoma		
FEATURES		FH Key Location/Qualifiers		
Source		FT source 1..904 /organism='Homo sapiens (human)'.		
		FT location/Qualifiers		
		1..904		
		/organism="Homo sapiens"		
		/mol_type="genomic DNA"		
		/db_xref="taxon:9606"		
ORIGIN				
Query Match		10.3%; Score 449; DB 6; Length 904;		
Best Local Similarity		100.0%; Pred. No. 1.4e+254;		
Matches 449; Conservative		0; Mismatches 0; Indels 0; Gaps 0;		
Oy		(160) GCTAACCGCTGTGATGAGCGTAGCCGATGCGATGAGATGCCAAGAAGCAAAACC	19000	
Df		52 GCTACCGCTGTGATGAGCGTAGCCGATGCGATGAGATGCCAAGAAGCAAAACC	111	
Oy		1901 TCACCGAGAGTGATGAGAAATAGCTGTCTGCGCAAAGAACCTGTCTTCAGATAGTTGAGC	1968	
Df		112 TCACCGAGAGTGATGAGAAATAGCTGTCTGCGCAAAGAACCTGTCTTCAGATAGTTGAGC	171	
Oy		1961 ATGCCATTCCGAGAGTGCGCACAGACCTGTATATATGTAACCTTTGTCTCAACATATGTTAT	20200	
Df		172 ATGCCATTCCGAGAGTGCGCACAGACCTGTATATATGTAACCTTTGTCTCAACATATGTTAT	231	
Oy		2021 CACTGCGATATATACCTTTTCATCTCCCTTGAATCTTTGTTTCACTATCTCGATTTAC	20800	
Df		232 CACTGCGATATATACCTTTTCATCTCCCTTGAATCTTTGTTTCACTATCTCGATTTAC	-291	

Qy	2081	AAAAACCTTTCATCGGCTAAATTGATGATGAGAGGGATGATGGATTTCTTTCCCT	2144
Db	292	AAAACCTTTTCATCGGCTAAATTGATGATGAGAGGGATGATGGATTTCTTTCCCT	351
Qy	2141	TTTTGGGAATGAGGCTCTCAAGCTAAAGCTATAGATGAGATTCAGAGTTTCAGG	2200
Db	352	TTTTGGGAATGAGGCTCTCAAGCTAAAGCTATAGATGAGATTCAGAGTTTCAGG	411
Qy	2201	GTCGTTTCTATACATTTGGCTATGTTAAAGGGGTAAAGGGCTCTCTTCAATAGATG	2260
Db	412	GTCGTTTCTATACATTTGGCTATGTTAAAGGGGTAAAGGGCTCTCTTCAATAGATG	471
Qy	2261	TGGAAGATGAAGCAGCCCTTCCTTTAGA	2289
Db	472	TGGAAGATGAAGCAGCCCTTCCTTTAGA	500
RESULT 15			
LOCUS	BD097758	904 bp	DNA linear
DEFINITION	Novel genes cloned in humaneuroblastoma and fragments thereof.		
ACCESSION	BD097758		
VERSION	BD097758.1	GI:22643332	
KEYWORDS	WO 0166719-A/58.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Bukaryolc, Metazoa, Chordata, Craniata, Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 904)		
AUTHORS	Novel genes cloned in humaneuroblastoma and fragments thereof		
TITLE	Patent: WO 0166719-A 58 13-SEP-2001;		
JOURNAL	CHIBA PREF, HISAMITSU PHARMACEUTICAL CO INC, AKIRA NAKAGAWARA		
COMMENT	OS Homo sapiens (human)		
	PN WO 0166719-A/58		
	PD -13-SEP-2001		
	PF 02-MAR-2001 WO 2001JP001629		
	PR 07-MAR-2000 JP 00P 159195		
	PI AKIRA NAKAGAWARA		
	PC C12N15/11, C1201/68, G01N33/53, G01N33/566		
	CC Novel genes cloned in humaneuroblastoma and fragments thereof		
	FN Key		
	FT source		
FEATURES	Location/Qualifiers		
source	1..904		
	/organism="Homo sapiens (human)"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	10.3%;	Score 449;	DB 6;
Best Local Similarity	100.0%;	Pred. No. 1,4e-254;	
Matches 449;	Conservative 0;	Mismatches 0;	Indels 0;
	Gaps 0;		
Qy	1841	GCTACCCGTGGTGGATGAGCGTACCGCATGCGATGAGATCCAGAAGACCAACCC	1900
Db	52	GCTACCCGTGGTGGATGAGCGTACCGCATGAGATGAGATCCAGAAGACCAACCC	111
Qy	1901	TCACCGGACGTGATGAGAAATAGCTGTGCTGGCAAGAACCTGTCTTCAGATAGTTGAGC	1960
Db	112	TCACCGGACGTGATGAGAAATAGCTGTGCTGGCAAGAACCTGTCTTCAGATAGTTGAGC	171
Qy	1961	ATGCGATTTCCGAGAGTGGCAGAGACCTGATATATGTATGATCCTTGTCTTCATATGTTAT	2020
Db	172	ATGCGATTTCCGAGAGTGGCAGAGACCTGATATATGTATGATCCTTGTCTTCATATGTTAT	231
Qy	2021	CACCTGCGTAAATACCCCTTTCATACCTCTGTGACCTTGTTCATTAATCTGTATTCAC	2080
Db	232	CACCTGCGTAAATACCCCTTTCATACCTCTGTGACCTTGTTCATTAATCTGTATTCAC	291
Qy	2081	AAAAACCTTTCATCGGCTAAATTGATGATGAGAGGGATGATGGATTTCTTTCCCT	2140

Db	292	AAAAACCTTTTCATTCCGCTAAATTGTGAGTTATGGAGGTGATTGGAAATTTCTTTCCCT	351
QY	2141	TTTTTGGGAAATGGGCTCTCAAGCTAAAGCTATAGATGGCAGATTCAAGAGTTTCAAGG	2200
Db	352	TTTTTGGGAAATGGGCTCTCAAGCTAAAGCTATAGATGGCAGATTCAAGAGTTTCAAGG	411
QY	2201	GTCGTGTTCTATATACATTGGCTATGTAAAGGGGTAAAGGGCTCTTTCATTATACATG	2266
Db	412	GTCGTGTTCTATATACATTGGCTATGTAAAGGGGTAAAGGGCTCTTTCATTATACATG	471
QY	2261	TGGAAGATGAAGCAGCCCTTCCTTTAGA	2289
Db	472	TGGAAGATGAAGCAGCCCTTCCTTTAGA	500
RESULT 16			
LOCUS	CQ706413	600 bp	DNA
DEFINITION	Sequence 51339 from Patent WO02070737.	linear	PAT 03-FEB-2002
ACCESSION	CQ706413		
VERSION	CQ706413.1	GI:42267182	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1		
AUTHORS	Liew,C.C., Marshall,W.E. and Zhang,H.		
TITLE	Compositions and methods relating to osteoarthritis		
JOURNAL	Patent: WO 02070737-A 51339 12-SEP-2002;		
FEATURES	Chondrogene Inc. (CA)		
SOURCE	Location/Qualifiers		
	1..600		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match			
Best Local Similarity 100.0%; Score 436; DB 6; Length 600;			
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	3548	ACACCTTAAGAGAGATTTCTAGAAAACTGGGCCAGATTTCTTTGTTCTCATCATY	3607
Db	32	ACACCTTAAGAGAGATTTCTAGAAAACTGGGCCAGATTTCTTTGTTCTCATCATY	91
QY	3608	TTAATGAGGAGGCGTTCAGTTTCTACCTTACCTTACATGATATTTCTTGTAAGT	3667
Db	92	TTAATGAGGAGGCGTTCAGTTTCTACCTTACCTTACATGATATTTCTTGTAAGT	151
QY	3668	GTCGCAAAAAGAAAAAGACCAATCAGTGTCTGTGACTTTGTTCTTGATCCCTCAGT	3727
Db	152	GTCGCAAAAAGAAAAAGACCAATCAGTGTCTGTGACTTTGTTCTTGATCCCTCAGT	211
QY	3728	TCTTCTTGATTCAGCATGTGTGCGGTTCTTAATTTGGGTATGAGTTAGCAATTTAAC	3787
Db	212	TCTTCTTGATTCAGCATGTGTGCGGTTCTTAATTTGGGTATGAGTTAGCAATTTAAC	271
QY	3788	CATTGTGTTTGTGCCCTACCCAGGGGATCTCCCACTTTCTGACTTGAAGTAGACGAGAA	3844
Db	272	CATTGTGTTTGTGCCCTACCCAGGGGATCTCCCACTTTCTGACTTGAAGTAGACGAGAA	331
QY	3848	GAATCCACGAGGTGCTATCTGGCCAGATTAAGTATTCATTTCTTGTTCTCCCTC	3907
Db	332	GAATCCACGAGGTGCTATCTGGCCAGATTAAGTATTCATTTCTTGTTCTCCCTC	391
QY	3908	TCCCTGAGGACCTCTTAATTTATTTGTCCTCTTCTAGATTAAATTCCTTGATTTGAC	3967
Db	392	TCCCTGAGGACCTCTTAATTTATTTGTCCTCTTCTAGATTAAATTCCTTGATTTGAC	451
QY	3968	TTTGTGAGAGGAGG 3983	

Db	452	TTTGTTGAGGAAGAGG	467
RESULT 17			
LOCUS	BD018137/c	893 bp	DNA linear PAT 27-AUG-2002
DEFINITION	Novel gene and novel gene fragment cloned in human neuroblastoma.		
ACCESSION	BD018137		
VERSION	BD018137.1	GI:22559313	
KEYWORDS	JP 2001245671-A/375.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 893)		
TITLE	Nakagawara, A.		
JOURNAL	Novel gene and novel gene fragment cloned in human neuroblastoma Patent: JP 2001245671-A 375 11-SEP-2001;		
COMMENT	CHTBA PRFP, HISAMITSU PHARMACEUTICAL CO INC		
	OS Homo sapiens (human)		
	PN JP 2001245671-A/375		
	PD 11-SEP-2001		
	PF 07-MAR-2000 JP 2000159195		
	PI AKIRA NAKAGAWARA		
	PC C12N15/09, C12Q1/68, G01N33/53, G01N33/5666 ((C12Q1/68, C12R1/91),		
	PC C12N15/00		
	CC Novel gene and novel gene fragment cloned in human CC neuroblastoma		
FEATURES			
Source	Location/Qualifiers		
	1..893		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	9.8%; Score 427; DB 6; Length 893;		
Best Local Similarity	100.0%; Pred. No. 1.7e-241;		
Matches 427, Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
OY	3912	TGAGAGCTCTTATTTTATTTGTCCCCCTCTCTAGTTAATCTCTTGATTTGACTTG	39717
Db	468	TGAGAGCTCTTATTTTATTTGTCCCCCTCTCTAGTTAATCTCTTGATTTGACTTG	409
OY	3972	TTGAGAGGAGGTGGACAGTATGATTTGCAAAAGTCCAAAGTCAAAATTAACAGTGTGA	4031
Db	408	TTGAGAGGAGGTGGACAGTATGATTTGCAAAAGTCCAAAGTCAAAATTAACAGTGTGA	349
OY	4032	GAGTGTGGGGGAAATTAATGCTTATTTTCCCTACATGGGATCAACACTGTGAATTC	4091
Db	348	GAGTGTGGGGGAAATTAATGCTTATTTTCCCTACATGGGATCAACACTGTGAATTC	289
OY	4092	ATCTTCACTGAAGCCCTGCAAGTTCCTTAAACATAGTGTGTTTCTTTTAAACA	4155
Db	288	ATCTTCACTGAAGCCCTGCAAGTTCCTTAAACATAGTGTGTTTCTTTTAAACA	229
OY	4152	AGTTAAGCTAGTGTAAATTAATTAATAAAAAAATGCTGCTGCTACTGAGCTTGT	4211
Db	228	AGTTAAGCTAGTGTAAATTAATTAATAAAAAAATGCTGCTGCTACTGAGCTTGT	169
OY	4212	TTATGCCATTTCATATTTGTTGTCGTGTGTGAATTCATACTTTGATACATTTCTGA	4277
Db	168	TTATGCCATTTCATATTTGTTGTCGTGTGTGAATTCATACTTTGATACATTTCTGA	109
OY	4272	TGTGTAATTTGTTGCTTGTAAATATCTTATTAAGATTCATTTGTAATTAACATAT	4333
Db	108	TGTGTAATTTGTTGCTTGTAAATATCTTATTAAGATTCATTTGTAATTAACATAT	49
OY	4332	GTGGCTG 4338	
Db	48	GTGGCTG 42	

```

RESULT 18
LOCUS      BD098075/c      893 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION Novel genes cloned in humanneuroblastoma and fragments thereof.
ACCESSION  BD098075
VERSION     BD098075.1 GI:22643649
KEYWORDS    WO 0166719-A/375.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE       Nakagawara, A.
JOURNAL     Novel genes cloned in humanneuroblastoma and fragments thereof
            Patent: WO 0166719-A 375 13-SEP-2001;
            CHIBA PREF,HISAITISU PHARMACEUTICAL CO INC,AKIRA NAKAGAWARA
COMMENT     OS Homo sapiens (human)
            PN WO 0166719-A/375
            PD 13-SEP-2001
            PR 02-MAR-2000 WO 2001JP001629
            PR 07-MAR-2000 JP 00P 159195
            PI AKIRA NAKAGAWARA
            PC C12N15/11,C12Q1/68,G01N33/53,G01N33/566
            CC Novel genes cloned in humanneuroblastoma and fragments thereof
            FH Key
            FT source      1. 893
                        Location/Qualifiers
                        .
                        location="Homo sapiens (human)".
FEATURES
    source      1. 893
                Location/Qualifiers
                .
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
ORIGIN
Query Match      9.8%; Score 427; DB 6; Length 893;
Best Local Similarity 100.0%; Pred. No. 1,7e-241;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      3912 TGAAGACCTTATTTATTTATGTCCTTCTAGGTTAATTCCTTGAATTGACTTTG 3971
DB      448 TGAAGACCTTATTTATTTATGTCCTTCTAGGTTAATTCCTTGAATTGACTTTG 409
QY      3972 TTGAGAGAGAGCTTGACAGTATTTGCAAGTTCAGTGCAGTATTCAGTGTGTA 4031
DB      408 TTGAGAGAGAGCTTGACAGTATTTGCAAGTTCAGTGCAGTATTCAGTGTGTA 349
QY      4032 GAGTGTGGGGGAAAAATTAGTCTTATTTTCCCTACATGGGATPACAACACTGTGAATTC 4091
DB      348 GAGTGTGGGGGAAAAATTAGTCTTATTTTCCCTACATGGGATPACAACACTGTGAATTC 289
QY      4092 ATCTTCACTGAAGCCCTCAGTCTCTTAAACATAGTGTGTTGTTTCTTTAAACA 4151
DB      288 ATCTTCACTGAAGCCCTCAGTCTCTTAAACATAGTGTGTTGTTTCTTTAAACA 229
QY      4152 AGTTTAAAGCTAGTCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4211
DB      228 AGTTTAAAGCTAGTCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 169
QY      4212 TTATGCCATTTTCAATATTTGTTGCTGTGTGTAATTCATATCACTTTGATTCATTTCTGA 4271
DB      168 TTATGCCATTTTCAATATTTGTTGCTGTGTGTAATTCATATCACTTTGATTCATTTCTGA 109
QY      4272 TGTGTAAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 4331
DB      108 TGTGTAAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 4331
QY      4332 GTGGCTG 4338
DB      48 GTGGCTG 42

```

```

RESULT 19
LOCUS      BV167952      580 bp      DNA      linear      STS 10-JUN-2004
DEFINITION segm689 Human DNA (Sequenom) Homo sapiens STS genomic, sequence
            tagged site.
ACCESSION  BV167952
VERSION     BV167952.1 GI:48001621
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE       Nelson,C.R., Marnellos,G., Kammerer,S., Hoyal,C.R., Shi,M.M.,
            Cantor,C.R. and Braun,A.
            Large-Scale Validation of Single Nucleotide Polymorphisms in Gene
            Regions
            Genome Res. (2004) In press
JOURNAL     Contact: Andreas Braun
            Pharmaceuticals division
            Sequenom, Inc.
            3595 John Hopkins Court, San Diego, CA 92121, USA
            Tel: 18582029018
            Fax: 18582029020
            Email: abraun@sequenom.com
            Primer A: No primer sequence submitted
            Primer B: No primer sequence submitted
            STS size: 580.
FEATURES
    source      1. 580
                Location/Qualifiers
                .
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /clone_lib="Human DNA (Sequenom)"
ORIGIN
Query Match      9.7%; Score 422; DB 11; Length 580;
Best Local Similarity 99.6%; Pred. No. 1.5e-238;
Matches 522; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      107 GGCCGGGGCCCCCGCGGCGGCAATCCTGAGCAGCACTGAGCTACAGAGGGGCTGGGC 166
DB      1 GGCCGGGGCCCCCGCGGCGGCAATCCTGAGCAGCACTGAGCTACAGAGGGGCTGGGC 60
QY      167 GCTGGGGGGGCGGAGAGCAGGCTGGGGGGGCGAAGCCCACTTCTCCCGGCAACG 226
DB      61 GCTGGGGGGGCGGAGAGCAGGCTGGGGGGGCGAAGCCCACTTCTCCCGGCAACG 120
QY      227 TANGCTCAAGAGACCGGGGCGGCGGCTGCTCTCCAGGGGGGAGCCCCGCGCTTC 286
DB      121 TANGCTCAAGAGACCGGGGCGGCGGCTGCTCTCCAGGGGGGAGCCCCGCGCTTC 180
QY      287 CCCGGCGGCTGGGGCGGCAAGGGCCGAGGCTTTTAACTCCCGGCGGAGCGGCGGCGG 346
DB      181 CCCGGCGGCTGGGGCGGCAAGGGCCGAGGCTTTTAACTCCCGGCGGAGCGGCGGCGG 240
QY      347 GCAAGCAGAGAGAGCTGGGGCGGCTTCCGCTTCCCTGTCTCCGCGCGGCAACAA 406
DB      241 GCAAGCAGAGAGAGCTGGGGCGGCTTCCGCTTCCCTGTCTCCGCGCGGCAACAA 300
QY      407 GCAAGCAGAGAGCTGGGGCGGAGCTTCCGCTTCCGCTTCCGCGCGGCAACAA 466
DB      301 GCAAGCAGAGAGCTGGGGCGGAGCTTCCGCTTCCGCTTCCGCGCGGCAACAA 360
QY      467 GTATCAAGCGGCTGCGGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCG 526
DB      361 GTATCAAGCGGCTGCGGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCG 420
QY      527 GCCCTGGGCTGGGAGCAAGGGTGGGGCGGCTTCCCGGCTGGCAACGCGCTGGACCGGCG 586
DB      421 GCCCTGGGCTGGGAGCAAGGGTGGGGCGGCTTCCCGGCTGGCAACGCGCTGGACCGGCG 480

```

QY	587	GGGAGCCCCC	CACCTACTCCTG	CGCCGGGCGG	CGACACCTC	GGGCG	630
Db	481	GGGAGCCCCC <td>CACCTACTCCTG <td>CGCCGGGCGG <td>CGACACCTC <td>GGGCG</td> <td>524</td> </td></td></td>	CACCTACTCCTG <td>CGCCGGGCGG <td>CGACACCTC <td>GGGCG</td> <td>524</td> </td></td>	CGCCGGGCGG <td>CGACACCTC <td>GGGCG</td> <td>524</td> </td>	CGACACCTC <td>GGGCG</td> <td>524</td>	GGGCG	524
RESULT 20							
LOCUS	BV178745		1092 bp	DNA	linear	STS 10-JUN-2004	
DEFINITION	BV178745		sgmm102348	Human DNA (Sequenom)	Homo sapiens STS genomic, sequence	tagged site.	
ACCESSION	BV178745						
VERSION	BV178745.1		GI:48015232				
KEYWORDS	STS.						
SOURCE	Homo sapiens						
ORGANISM	Homo sapiens (human)						
REFERENCE	Bukatyk; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1092)						
AUTHORS	Nelson, R.M., Marinello, G., Kammerer, S., Hoyal, C.R., Shi, M.M., Cantor, C.R. and Braun, A.						
TITLE	Large-Scale Validation of Single Nucleotide Polymorphisms in Gene Regions						
JOURNAL	Genome Res. (2004)						
COMMENT	In press						

Email: abraun@sequenom.com
 Primer A: No primer sequence submitted
 Primer B: No primer sequence submitted
 STS size: 1092.

```
location/Qualifiers
1..1092
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_l1b="Human DNA (Sequencuon)"
1..>1092
```

Query Match	9.7%	Score 422	DB 11	Length 1092
Best Local Similarity	99.6%	Pred. No. 1.7e-238		
Matches 522	Conservative 0	Mismatches 2	Indels 0	Gaps 0

Query Match	9.7%	Score 422	DB 11	Length 1092
Best Local Similarity	99.6%	Pred. No. 1.7e-238		
Matches 522	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY	107	GGCCGCGGCCCCCTGCGCGGCGCAATCTCGAGAGGCACTGGACTTACGAGCGGGCTGCGGC	166	
Db	1	GGCCGCGGCCCCCTGCGCGGCGCAATCTCGAGAGGCACTGGACTTACGAGCGGGCTGCGGC	60	
QY	167	GCTGGGCGGGGCGCGAGGAGCGAGCCCTGGGGGCGGCGGAAGCCCACTTCTCCCGGCAACG	226	
Db	61	GCTGGGCGGGGCGCGAGGAGCGAGCCCTGGGGGCGGCGGAAGCCCACTTCTCCCGGCAACG	120	
QY	227	TAAGTCAGGAGCCGGGGGCCCCGCTGAGCTCTCCCAAGGGCGGGAGGCCCGCGCCTTTC	286	
Db	121	TAAGTCAGGAGCCGGGGGCCCCGCTGAGCTCTCCCAAGGGCGGGAGGCCCGCGCCTTTC	180	
QY	287	CCCCGCGCGCTGCGGCGGCAAGGGGCGGGGGCTTTGTTACTCCCGGGCGGGGCGGCCCGGG	346	
Db	181	CCCCGCGCGCTGCGGCGGCAAGGGGCGGGGGCTTTGTTACTCCCGGGCGGGGCGGCCCGGG	240	
QY	347	GCAGCAGAGAGAGAGACTGGGGCGGTTCCGATGCGCTTTCGCGCCCGCGCACCA	406	
Db	241	GCAGCAGAGAGAGAGACTGGGGCGGTTCCGATGCGCTTTCGCGCCCGCGCACCA	300	
QY	407	GCAAGCCGGCAATTGGGGGGGAGGACTTGCACGACCGAGGCCGAGCTCAGGCCCGGCCCA	466	
Db	301	GCAAGCCGGCAATTGGGGGGGAGGACTTGCACGACCGAGGCCGAGCTCAGGCCCGGCCCA	360	
QY	467	GTATCAGCGGCTGCTCCCATTTACAGACGGGCTCTCTGCTGGCGGGCGGCCCAAGAGCTTAC	526	

Db 361 GTATCAAGCGGTGCTGCCATTCAAGAGGGCTCTCTCTGTGGCGGGCGGCACCAAGAGCTTAC 420

Qy 527 GCCCTGGGCTTGAGGACAAGGGTGGGCGGCTCCCTCGTGCACCGGCTTCGACCCGGC 586

Db 421 GCCCTGGGCTTGAGGACAAGGGTGGGCGGCTCCCTCGTGCACCGGCTTCGACCCGGC 480

Qy 587 GGGACCCCGACCACTACTCTCTGCCCGGGCGGCACCCCTCGGGC 630

Db 481 GGGACCCCGACCACTACTCTCTGCCCGGGCGGCACCCCTCGGGC 524

RESULT 21					
LOCUS	C0713127	451 bp	DNA	linear	PAT 03-FEB-2004
DEFINITION	Sequence 58053 from Patent WO02070737.				
ACCESSION	C0713127				
VERSION	C0713127.1	GI:42273984			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE	1. Liew C.C., Marshall, W.B. and Zhang H. Compositions and methods relating to osteoarthritis Patent: WO 02070737-A 58053 12-SEP-2002; Chondrogene Inc. (CA)				
AUTHORS	Location/Qualifiers				
TITLE	1. .451				
JOURNAL					
FEATURES					
source					

```

/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

```

Query Match	9.4%	Score 410	DB 6	length 451
Best Local Similarity	100.0%	Pred. No. 2e231		
Matches 410; Conservative	0	Mismatches	0	Gaps 0

QY	3875	TTTAAAGTAGATTCAATTTCCCTGGATCTCCCTCCCGGAGAGCCCTTAATTTTATTTGTC	3934
Db	7	TTTAAAGTAGATTCTAATTTCCCTGGATCTCCCTCCCGGAGAGCCCTTAATTTTATTTGTC	66
QY	3935	CCCTCTTCTAGATTAAATTTCTCCTTTGATTTGACTTTGTGTAGAAAGAGCTTGACAGTAG	3994
Db	67	CCCTCTTCTAGATTAAATTTCTCCTTTGATTTGACTTTGTGTAGAAAGAGAGTTGACAGTAG	126
QY	3995	ATTAGCAAAAGTTCCAGATGCAGAAATTTACAGTGTGTAGAGTGTGGGGGAAAATTAGTCT	4055
Db	127	ATTAGCAAAAGTTCCAGATGCAGAAATTTACAGTGTGTAGAGTGTGGGGGAAAATTAGTCT	186
QY	4055	TATTTTTCCTACACAGGAGTACAAACACGTGAAATTCATCTTCAACTAGAGCCCTGACG	4114
Db	187	TATTTTTCCTACACAGGAGTACAAACACTGTAAATTCATCTTCAACTAGAGCCCTGACG	246
QY	4115	TTCTCTCTAAACATAGTGTGTGTCTTTTCTTTTAACAAAAGTTTAAAGTATGTATAAAT	4174
Db	247	TTCTCTCTAAACATAGTGTGTGTCTTTTCTTTTAACAAAAGTTTAAAGTATGTATAAAT	306
QY	4175	TAAAAAAAATTCGTTGCTGTCTACTACCTTCAGCTTGTGTTTATGCCCCATTTCAATATGTTGT	4233
Db	307	TAAAAAAAATTCGTTGCTGTCTACTACCTTCAGCTTGTGTTTATGCCCCATTTCAATATGTTGT	366
QY	4235	CTGTGTTGTAAATTCATAACTTTTGTATACAAATTTCTGATGTGTAAATTTGG	4284
Db	367	CTGTGTTGTAAATTCATAACTTTTGTATACAAATTTCTGATGTGTAAATTTGG	416

RESULT 22	BD018136	BD018136	903 bp	DNA	linear	PAT 27-AUG-2007
LOCUS	BD018136					
DEFINITION	Novel gene and novel gene fragment cloned in human neuroblastoma.					
ACCESSION	BD018136					

VERSION BD018136.1 GI:22559312
KEYWORDS JP 2001245671-A/374.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 903)
JOURNAL Nakagawara, A.
Novel gene and novel gene fragment cloned in human neuroblastoma
Patent: JP 2001245671-A 374 11-SEP-2001;
CHIBA PREF,HISAMITSU PHARMACEUTICAL CO INC
COMMENT OS Homo sapiens (human)
PN JP 2001245671-A/374
PD 11-SEP-2001
PF 07-MAR-2000 JP 2000159195
PI AKIRA NAKAGAWARA
PC C12N15/09,C12Q1/68,G01N33/53,G01N33/566/(C12Q1/68,C12R1.91),
CC C12N15/00
Novel gene and novel gene fragment cloned in human CC
neuroblastoma
FH Key Location/Qualifiers
FT source 1..903
FT /organism='Homo sapiens (human)'.
FEATURES
source 1..903
Location/Qualifiers
1..903
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Query Match 7.0%; Score 306; DB 6; Length 903;
Best Local Similarity 100.0%; Pred. No. 1.7e-169;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1877 GGAGATCCAGAGAAGCAAAACACCTCACCGAGCTGTAGGAATAGCTGTGGCAAGA 1936
DB 117 GGAGATCCAGAGAAGCAAAACACCTCACCGAGCTGTAGGAATAGCTGTGGCAAGA 176
QY 1937 ACCCTGCTTCAGATAGTTGTAGATGCCATTCCCGAGAGTGGAGAGACCTGTATATGT 1996
DB 177 ACCCTGCTTCAGATAGTTGTAGATGCCATTCCCGAGAGTGGAGAGACCTGTATATGT 236
QY 1997 GACCTTGTCTCACAATATGTATCACTCGCTGATTAATACCTTTCACTTCCCTTACT 2056
DB 237 GACCTTGTCTCACAATATGTATCACTCGCTGATTAATACCTTTCACTTCCCTTACT 296
QY 2057 TTGTTTCATTACTCTGTATTCACAAAACCTTTCACTTGGCTAATTTGAGTTATGA 2116
DB 297 TTGTTTCATTACTCTGTATTCACAAAACCTTTCACTTGGCTAATTTGAGTTATGA 356
QY 2117 GGGTGAATGGGATTTCTTTCCCTTTTGGGAAATGGGCTCCAACTAAAGCTATAGG 2176
DB 357 GGGTGAATGGGATTTCTTTCCCTTTTGGGAAATGGGCTCCAACTAAAGCTATAGG 416
QY 2177 ATGGCA 2182
DB 417 ATGGCA 422
RESULT 23
BD098074 903 bp DNA linear PAT 27-AUG-2002
LOCUS BD098074
DEFINITION Novel genes cloned in humanneuroblastoma and fragments thereof.
ACCESSION BD098074.1 GI:22643648
VERSION BD098074.1
KEYWORDS WO 0166719-A/374.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 903)
JOURNAL Nakagawara, A.
Novel genes cloned in humanneuroblastoma and fragments thereof

JOURNAL Patent: WO 0166719-A 374 13-SEP-2001;
COMMENT CHIBA PREF,HISAMITSU PHARMACEUTICAL CO INC,AKIRA NAKAGAWARA
OS Homo sapiens (human)
PN WO 0166719-A/374
PD 13-SEP-2001
PF 02-MAR-2001 WO 2001JP001629
PI AKIRA NAKAGAWARA
PC C12N15/11,C12Q1/68,G01N33/53,G01N33/566
CC Novel genes cloned in humanneuroblastoma and fragments thereof
FH Key Location/Qualifiers
FT source 1..903
FT /organism='Homo sapiens (human)'.
FEATURES
source 1..903
Location/Qualifiers
1..903
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Query Match 7.0%; Score 306; DB 6; Length 903;
Best Local Similarity 100.0%; Pred. No. 1.7e-169;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1877 GGAGATCCAGAGAAGCAAAACACCTCACCGAGCTGTAGGAATAGCTGTGGCAAGA 1936
DB 117 GGAGATCCAGAGAAGCAAAACACCTCACCGAGCTGTAGGAATAGCTGTGGCAAGA 176
QY 1937 ACCCTGCTTCAGATAGTTGTAGATGCCATTCCCGAGAGTGGAGAGACCTGTATATGT 1996
DB 177 ACCCTGCTTCAGATAGTTGTAGATGCCATTCCCGAGAGTGGAGAGACCTGTATATGT 236
QY 1997 GACCTTGTCTCACAATATGTATCACTCGCTGATTAATACCTTTCACTTCCCTTACT 2056
DB 237 GACCTTGTCTCACAATATGTATCACTCGCTGATTAATACCTTTCACTTCCCTTACT 296
QY 2057 TTGTTTCATTACTCTGTATTCACAAAACCTTTCACTTGGCTAATTTGAGTTATGA 2116
DB 297 TTGTTTCATTACTCTGTATTCACAAAACCTTTCACTTGGCTAATTTGAGTTATGA 356
QY 2117 GGGTGAATGGGATTTCTTTCCCTTTTGGGAAATGGGCTCCAACTAAAGCTATAGG 2176
DB 357 GGGTGAATGGGATTTCTTTCCCTTTTGGGAAATGGGCTCCAACTAAAGCTATAGG 416
QY 2177 ATGGCA 2182
DB 417 ATGGCA 422
RESULT 24
CQ663907 400 bp DNA linear PAT 03-FEB-2004
LOCUS CQ663907
DEFINITION Sequence 8833 from Patent WO02070737.
ACCESSION CQ663907
VERSION CQ663907.1 GI:42143690
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 903)
JOURNAL Chondrogene Inc. (CA)
Patent: WO 02070737-A 8833 12-SEP-2002;
Chondrogene Inc. (CA)
FEATURES
source 1..400
Location/Qualifiers
1..400
/organism='Homo sapiens'
/mol_type='unassigned DNA'
/db_xref='taxon:9606'
ORIGIN
Query Match 6.4%; Score 281; DB 6; Length 400;

Best Local Similarity 100.0%; Pred. No. 1.1e-154;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3949 AATTCTCCTTTGATTGACTTTGTTGAGAGAGGTTGACAGTAGATTAGCAAAAGTTCC 4008
DB 69 AATTCTCCTTTGATTGACTTTGTTGAGAGAGGTTGACAGTAGATTAGCAAAAGTTCC 128
QY 4009 AAGTGCAGAAATTCAGATGTGTAGAGTGTGGGGGAGAAATTAATTAATTTTTCCTTACA 4068
DB 129 AAGTGCAGAAATTCAGATGTGTAGAGTGTGGGGGAGAAATTAATTAATTTTTCCTTACA 188
QY 4069 TGGGATACACACTGTGAATTCATCTTCAACTGAAAGCCCTGACGTTCTCTTAAACAT 4128
DB 189 TGGGATACACACTGTGAATTCATCTTCAACTGAAAGCCCTGACGTTCTCTTAAACAT 248
QY 4129 AGTTGTTGTTTTCCTTAAACAAAGTTAAGCTAGTGTAAATTAATTAATTAATTTGCT 4188
DB 249 AGTTGTTGTTTTCCTTAAACAAAGTTAAGCTAGTGTAAATTAATTAATTAATTTGCT 308
QY 4189 TGTCTGTCTACTTACGCTTTGTTTATGCCCATTTCAATTT 4229
DB 309 TGTCTGTCTACTTACGCTTTGTTTATGCCCATTTCAATTT 349

RESULT 25

G22752 240 bp DNA linear STS 31-MAY-1996
LOCUS human STS WI-17575, sequence tagged site.
DEFINITION G22752
ACCESSION G22752.1 GI:1343078
VERSION
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 240)
AUTHORS Hudson.T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STS
JOURNAL Unpublished (1995)
COMMENT

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: AAATATCAGATAGCACTCCCA
Primer B: ACTGTGTTTCAACAATAGTCTCG
STS size: 150
PCR Profile:

Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:

Protocol:
Template: 10 ng
Primer: each 5 pm
dNTPs: each 4 mM
Tag Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

Derived from dbEST (genbank accession R78466).

FEATURES
source Location/Qualifiers
1..240
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="319.5 cR from top of Chr17 linkage group"

STSS
primer_bind 1..23
primer_bind complement(127..150)

Query Match 5.5%; Score 240; DB 11; Length 240;
Best Local Similarity 100.0%; Pred. No. 2.4e-130; Indels 0; Gaps 0;
Matches 240; Conservative 0; Mismatches 0;

QY 2943 AAAATATCAGATAGCACTCCAGGCACTTTGTTGCTCAGTGAATCCCTTAATTA 3002
DB 1 AAAATATCAGATAGCACTCCAGGCACTTTGTTGCTCAGTGAATCCCTTAATTA 60
QY 3003 TCTGAAAGGAAATATAGAGCAAGCACTGTGCTCAATATATAGAAATTCCTTTCTTT 3062
DB 61 TCTGAAAGGAAATATAGAGCAAGCACTGTGCTCAATATATAGAAATTCCTTTCTTT 120
QY 3063 AGTCTTCAAGACTATGTGTGAAGCAAGTATGCTATCTCTAGAAAGTATGAGGCT 3122
DB 121 AGTCTTCAAGACTATGTGTGAAGCAAGTATGCTATCTCTAGAAAGTATGAGGCT 180
QY 3123 TTTATCCTTAAAGAAATATGTCCTCCAGATTATTAACCTTTTATGAGAGAACCAAGGT 3182
DB 181 TTTATCCTTAAAGAAATATGTCCTCCAGATTATTAACCTTTTATGAGAGAACCAAGGT 240

RESULT 26
LOCUS C0698720 245 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 43646 from Patent WO02070737.
ACCESSION C0698720
VERSION C0698720.1 GI:42254016
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE Liew.C.C., Marshall.W.E. and Zhang,H.
JOURNAL Compositions and methods relating to osteoarthritis
Patent: WO 02070737-A 43646 12-SEP-2002;
Chondrogene Inc. (CA)

FEATURES
source Location/Qualifiers
1..245
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 5.5%; Score 239; DB 6; Length 245;
Best Local Similarity 100.0%; Pred. No. 9.4e-130; Indels 0; Gaps 0;
Matches 239; Conservative 0; Mismatches 0;

QY 2471 AGATACTGTCTGAAAGATAGTTGTCATTTGAAACTATTCGTGATACAGTCATGTGG 2530
DB 1 AGATACTGTCTGAAAGATAGTTGTCATTTGAAACTATTCGTGATACAGTCATGTGG 60
QY 2531 GAAAGATGTTGGCTGTGATTTATTTTTCAGTTANAGATACAATTTCTTACTGCTC 2590
DB 61 GAAAGATGTTGGCTGTGATTTATTTTTCAGTTATGATTAACAATTTCTTACTGCTC 120
QY 2591 AAAAACCAAAATCTTTGAAAAAGAAAGTGGGATGTTTCAAGAACCAAGTTACAGCT 2650
DB 121 AAAAACCAAAATCTTTGAAAAAGAAAGTGGGATGTTTCAAGAACCAAGTTACAGCT 180
QY 2651 GTTAAACCAAGCACTTAGTATTTTGGATGTCAGCAAAACCTGTATAATATGTCCTTGT 2709

RESULT	27
BC058629	
LOCUS	
DEFINITION	Mus musculus RIKEN cDNA 4121402D02 gene, mRNA (cDNA clone MGC:64809 IMAGE:6402215), complete cds.
ACCESSION	BC058629
VERSION	BC058629.1 GI:35193129
KEYWORDS	MG.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 2933)
AUTHORS	Strauberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shamen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhut,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J.F., Heien,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S., Canciani,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Boeck,S.A., McEwan,P.J., McKernan,K.C., Malek,U.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.U., Hulyk,S.W., Villallon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahney,J., Helton,E., Kettelman,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smalins,D.E., Scherch,A., Schein,J.E., Jones,S.J. and Marr,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED	12477932
REFERENCE	2 (bases 1 to 2933)
AUTHORS	Strauberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (22-SRP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: gcgabs-remail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library Preparation: The Bento Soares, University of Iowa DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr. Thomas L. Casavant. Web site: http://genome.uiowa.edu Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A., Fishler,K., Kepel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K., Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J., Casavant,T., Soares,M.B.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Series: Plate: Row: Column: 0 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21311987. Location/Qualifiers 1..2933 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6" /db_xref="taxon:10090" /clone="MGC:64809 IMAGE:6402215" /tissue_type="Mouse, Brain 12.5 dpc"

gene

1. 2933

/gene="4121402D02Rik"

/note="synonym: 281001F12Rik"

/db_xref="Locustid:74026"

/db_xref="MG1:1921276"

63. .1913

/gene="4121402D02Rik"

/codon_start=1

/product="RIKEN CDNA 4121402D02"

/protein_id="AAH58629.1"

/db_xref="GI:35193130"

/db_xref="Locustid:74026"

/db_xref="MG1:1921276"

/translation="MTTRAAVFRAAAAPAGNPEORLDYERAAALGPEDESGAAEAH
FLRRRLKEPPLASOGSGSPSPAGCGGGRGLPAGAPAOEESWGSVPL
LPORPTKOGAGIGERVAAGAGCSPPKYOAVLPLOTGSIVAAAEPEPMGDKGG
AAPPAATSDPACPPPLPLGPPPLAFTATAGTLLASBGMKSIKSPJLOGGSGSGAS
SQACAKQIILLQDLIEQQOQLOAKETIIEKSRDTLLIATEMERKMOIVKRD
NEBERHLLQGYPEREEREALESEKI KLEQPELCEISOALPSKPSGSGSGKHKRK
TTPCNERKTPVKKLLAPESFKVTKTPKXSIPIKEBPGSISEYCKKELISOETPEP
RSGVDNPEPLSTPOKSGTPKPKAFSSSEMEDLPYSTMTLCRMHOPPSPLPLBE
SSPKBETVAKCPSSVAGETSVLAAPSWRDHSVEPLRDNPSEDLLENDDSVSKR
HATLEDEKRRKRMWDITREQRIILQRLQIRMTKKGIQSEPEYTSFPPEPDVESL
LITPLPEVAAGRPPLPLAPQNELPULDRSCRIETIQKHTPHRCR"

CDS

ORIGIN	Query Match	5.0%; Score 219; DB 10; Length 2933;
	Best Local Similarity	100.0%; Pred. No. 1.2e-117; Mismatches 0; Gaps 0;
	Matches 219; Conservative	0; Indels 0;
QY	1338 GAAGATTGGCGTACCTTTCCACCACCAAGAAATGATTGTGTGCTGTGGACACAGCCTCCC	1397
Db	1329 GAAGATTGGCGTACCTTTCCACCACCAAGAAATGATTGTGTGCTGTGGACACAGCCTCCC	1388
QY	1398 CCATCACCGTTTACCAATTACGGGAAATCTCTTCCAAAGAGAGAGACTGTAGCAAGGTGT	1457
Db	1389 CCATCACCGTTTACCAATTACGGGAAATCTCTTCCAAAGAGAGAGACTGTAGCAAGGTGT	1448
QY	1458 CTGATGCGATCAAGTGTGTGAGAGAGAACTTCAGTCTGTTCCTTCTTGGAGAGGAC	1517
Db	1449 CTGATGCGATCAAGTGTGTGAGAGAGAACTTCAGTCTGTTCCTTCTTGGAGAGGAC	1508
QY	1518 CACTCAGTAGAGCCTCTAAGGAGACCCCAATCTTTCAGAC	1556
Db	1509 CACTCAGTAGAGCCTCTAAGGAGACCCCAATCTTTCAGAC	1547
RESULT 28		
LOCUS	CO687002	440 bp DNA linear PAT 03-FEB-2004
DEFINITION	Sequence 31928 from Patent WO02070737.	
ACCESSION	CO687002	
VERSION	CO687002.1	GI:42217066
KEYWORDS		
SOURCE		
ORGANISM	Homo sapiens (human)	
	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1	
AUTHORS	Liew,C.C., Marshall,W.E. and Zhang,H.	
TITLE	Compositions and methods relating to osteoarthritis	
JOURNAL	Patent: WO 02070737-A 31928 12-SEP-2002;	
	Chondrogene Inc. (CA)	
FEATURES	Location/Qualifiers	
SOURCE	1..440	
	/organism="Homo sapiens"	
	/mol_type="unassigned DNA"	
ORIGIN	/db_xref="taxon:9606"	

	Query Match	Similarity	4.9%	Score 214	DB 6	Length 440	
	Best Local	Similarity	100.0%	Pred. NO. 7.8e-115			
	Matches	214	Conservative	0	Mismatches	0	Indels
							Gaps 0
QY	1240	AAAAGCCCGGCTTTCAGTGGACA	CCCCCAAGA	CTTCCACTCCCAAAAGGACCCA	12399		
		5	AAAAGCCCGGCTTTCAGTGGACA	CCCCCAAGA	CTTCCACTCCCAAAAGGACCCA	64	
Db	1300	GCACCCATCCCAAGAGAAAGCCTTCTCAAGTGAATGAAGATTGGCGTACCTTTCCA	1355				
QY	65	GCACCCATCCCAAGAGAAAGCCTTCTCAAGTGAATGAAGATTGGCGTACCTTTCCA	124				
Db	1360	CCACAGAATATATTGTGTGCTTGGCCACAGCCTCCCCATCAACGTTACATTACGGG	1419				
QY	125	CCACAGAATATATTGTGTGCTTGGCCACAGCCTCCCCATCAACGTTACATTACGGG	184				
Db	1420	AATCCTCTCCCAAGAGAGAGATTTAGCAAG	1453				
QY	185	AATCCTCTCCCAAGAGAGAGATTTAGCAAG	218				

[illegible]

LOCUS	AR277392	542 bp	DNA	linear	PAT 10-APR-2003
DEFINITION	Sequence	1554 from patent US 6509448.			
ACCESSION	AR277392				
VERSION	AR277392.1	GI:29711039			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 542)				
TITLE	Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S., Carter,D., Reiter,M.W., Mannion,J., Fan,L. and Wang,A. Compositions and methods for the therapy and diagnosis of lung cancer				
JOURNAL	Patent: US 6509448-A 1554 21-JAN-2003;				
FEATURES	Location/Qualifiers				
source	1..542				
	/organism="unknown"				
	/mol_type="genomic DNA"				

Query Match	3.5%; Score 152; DB 6; Length 542;
Best Local Similarity	98.9%; Pred. No. 6,2e-78;
Matches 352; Conservative	0; Mismatches 4; Indels 0; Gaps 0;
QY	3395 TGTGNTGTGNTGTGNTGTGAGCTATGGATTTCATTGTGAACCTCCTTAGAG 3454
DB	66 TGTGTGTGTGTGTGTGTGTGTGGCTATGGGTTTCAATTGTAACTCCATCGTTAGAG 125
QY	3455 AGTGGGCTCTCTTAATPAGGGAACCTGCCTGTAAACTTTCATTGCGACGAGATGTAGAGAA 3514
DB	126 AGTGCGCTCTCTTAATPAGGGAACCTGCCTGTAAACTTCATTGCGACGAGATGTAGAGAA 185
QY	3515 ATAGACTTAATTCACATGAGGGGCTCATCTCACACCTTAAGGAGAGATTTCTAGAAA 3574
DB	186 ATAGACTTAATTCACATGAGGGGCTCATCTCACACCTTAAGGAGAGATTTCTAGAAA 245
QY	3575 AACTGGGCCAATTTCTTTGTTCTCCATCATTTTAATGTGGCAGGCTGTTCAGTTTTCT 3633
DB	246 AACTGGGCCAATTTCTTTGTTCTCCATCATTTTAATGTGGCAGGCTGTTCAGTTTTCT 305
QY	3635 TACTCTTAACCTATGTGATATTTCTTCGTAAAGGTCCAAAAAGAAAAAGAACCCATCAG 3694
DB	306 TACTCTTAACCTATMGATATTTCTTCGTAAAGGTCCAAAAAGAAAAAGAACCCATCAG 365
QY	3695 TGCTCTTGACTTTGTTCTTTTGATTCCTCAAGTTCTTCTTGATTTCAGCATGTGTC 3750
DB	366 TGCTCTTGACTTTGTTCTTTTGATTCCTCAAGTTCTTCTTGATTTCAGCATGTGTC 421
RESULT 31	
LOCUS	AR407667 542 bp DNA linear PAT 18-DEC-2003
DEFINITION	Sequence 1554 from patent US 6630574.
ACCESSION	AR407667
VERSION	AR407667.1 GI:40157478
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 542)
TITLE	Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S., Carter,D., Retter,M.W., Mannion,J. and Fan,L. Compositions and methods for the therapy and diagnosis of lung cancer
JOURNAL FEATURES	Patent: US 6630574-A 1554 07-OCT-2003; Location/Qualifiers 1..542 /organism="unknown" /mol_type="genomic DNA"
ORIGIN	
Query Match	3.5%; Score 152; DB 6; Length 542;

[illegible][illegible]

[illegible]

OY		4359	AAAAAAAAAAAAA	4372
Dd		172	AAAAAAAAAAAAAA	185
RESULT_36				
AL590963/c				
LOCUS				
DEFINITION		AL590963	196724 bp	DNA linear ROD 05-APR-2002
ACCESSION		AL590963		
VERSION		AL590963.11	G1:20068458	
KEYWORDS				
SOURCE		Mus musculus	(house mouse)	
ORGANISM				
REFERENCE		1		
AUTHORS		Lovell,J.		
TITLE		Direct Submission		
JOURNAL		Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humgier@sanger.ac.uk Clone requests: clonesrequest@sanger.ac.uk On Apr 7, 2002 this sequence version replaced g1:16304739. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted; all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw:', SWISSPROT; Tr:, TREMBL; Wp:, WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep/ RP23-395E10 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBacc3.6.		
FEATURES				
Source		Location/Qualifiers		
		1..196724		
	/organism="Mus musculus"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:10090"			
	/chromosome="11"			
	/clone="RP23-395E10"			
	/clone_idb="RPCI-23"			
ORIGIN				
	Query Match	2.9%; Score 127;	DB 10;	Length 196724;
	Best Local Similarity	100.0%;	Pred No. 1.3e-62;	
	Matches 127;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;
OY		1632	AGATGGATTTCGAGAGATCAGAGAACAAAGAATTTTA	CAGGCCTGCGAGTCAAGATG 1691
Dd		54277	AGATGGATTTCGAGAGATCAGAGAACAAAGAATTTTA	CAGGCCTGCGAGTCAAGATG 54218
OY		1692	TATATAAAGGAAGATTCGAGAACTCGAGCCGTGAGTTACTCATTTTTTCCCCTGAGCCA	1751
Dd		54217	TATATAAAGGAAGATTCGAGAACTCGAGCCGTGAGTTACTCATTTTTTCCCCTGAGCCA	54158
OY		1752	GATGATG 1758	
Dd		54157	GATGATG 54151	

FEATURES	source	Location/Qualifiers	
* 145222		148094: contig of 2873 bp in length	
* 148095		148194: gap of unknown length	
* 148195		183618: contig of 35424 bp in length.	
FEATURES	source	Location/Qualifiers	
misc_feature		1. 183618	
		/organism="Rattus norvegicus"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:10116"	
		/clone="CH230-278A19"	
		1. 1263	
		/note="wgs_contig"	
ORIGIN	Query Match	2.7%: Score 119; DB 2; Length 183618;	
	Best Local Similarity	100.0%; Pred. No. 7.4e-58;	
	Matches 119; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1632	AGATGGGATATTTCAGAGGATCAGGAAACAAAGATTTTACAGCAGCTGACGCTCAGATG 1691	
Db	58659	AGATGGGATATTTCAGAGGATCAGGAAACAAAGATTTTACAGCAGCTGACGCTCAGATG 58660	
Qy	1692	TATATAAAGAAAGAAATTCAGAAATCTGAGCCTGAGCTTACCTCATTTTCCCTGAGCC 1750	
Db	58599	TATATAAAGAAAGAAATTCAGAAATCTGAGCCTGAGCTTACCTCATTTTCCCTGAGCC 58541	
RESULT 39	AC119484/c	AC119484/c	
LOCUS	AC119484/c	AC119484/c	
DEFINITION	AC119484	230913 bp DNA linear HTG 19-NOV-2002	
ACCESSION	AC119484.5	GI:25095516	
VERSION	HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.		
KEYWORDS	Rattus norvegicus (Norway rat)		
SOURCE	Rattus norvegicus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 230913)		
AUTHORS	Munhy,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Albrooks,S., Amin,A., Angiano,D., Annelaebchi,V., Aoyagi,I., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barneshead,M., Benahmed,F., Bismail,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Butrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davis,M.,L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Diaper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Gatta,K., Garcia,A., Garner,T., Garza,M., Georgerogis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W., Gunaratne,P., Haaland,W., Hamli,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hlady,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Kapachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,K., Kwois,C., Krefc,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lounsbury,L., Louised,H., Loraado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mamoud,M., Malloy,K., Mangun,A., Mangun,B., Nagua,P., Martin,K., Martin,R., Martinez,E., Mawhiley,S., McLeod,M.P., McNeill,T.Z., Meenen,B., Milosavljevic,A., Miner,G., Muija,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naik,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwoketehem,O., Okwono,G., Olampunsaon,A., Pal,S., Parks,K., Paternack,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,		

Plazzo, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzer, M., Oltroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Rially, B., Reilly, M., Ren, Y., Router, M., Rose, R., Ruiz, S.-J., Sanders, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S.-J., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Swalek, A., Tabors, Z., Uemari, K., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, J., Valae, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Unpublished
2 (bases 1 to 230913)

REFERENCE
JOURNAL
TITLE
AUTHORS

Direct Submission
Submitted (27-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
JOURNAL
TITLE

Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 230913)

REFERENCE
AUTHORS
JOURNAL
TITLE

Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23269735.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atlas/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GVPH
Center clone name: CH230-38A24

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 218750 bases at least Q40
Consensus quality: 222810 bases at least Q30
Consensus quality: 224623 bases at least Q20
Estimated insert size: 234093; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 173284: contig of 173284 bp in length
* 173285 173384: gap of unknown length
* 173385 205324: contig of 31940 bp in length
* 205325 205424: gap of unknown length

Query Match 2.7%; Score 119; DB 2; Length 230913;
Best Local Similarity 100.0%; Pred. NC. 7.7e-58;
Matches 119; Conservative 0; Indels 0; Gaps 0;

RESULT_40	
BC055715	
LOCUS	
DEFINITION	
Mus musculus RIKEN CDNA IMAGE:614174), complete cds.	
BC055715	2087 bp
Mus musculus RIKEN CDNA IMAGE:614174), complete cds.	mRNA linear ROD 21-OCT-2003
	gene, mRNA (CDNA clone MGC:667878)

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
BC055715.1	GI:33585901	MGC	Mus musculus (house mouse)		
			Mus musculus		
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2087)		
				Struhsberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G.,	

TITLE	Journal, MEDLINE, PUBMED, AUTHOR, TITLE, JOURNAL
Altschul, S.F., Zeeberg, B., Butow, K.H., Schaefer, C.F., Bat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Dichtenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stepleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schaefer, T.E., Brownstein, M.J., Uedini, T.B., Toshiyuki, S., Caramici, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McEernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, P., Wexley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hilly, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Rane, J., Helton, E., Kettelman, M., Madan, A.C., Rodrigues, S., Sacher, A., Whiting, M., Madan, A., Young, A.C., Shcherbko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Girmwood, J., Schmitt, J., Myers, R.M., Buterfield, Y.S., Krzywinski, M.I., Skalka, U., Smalls, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marr, M.A., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
JOURNAL, MEDLINE, PUBMED, AUTHOR, TITLE, JOURNAL	2 (bases 1 to 2087) Straussberg, R. Direct Submission Submitted (01-AUG-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer

Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINI)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-engc.stanford.edu>
Contact: (Dickson, Mark) mcd@pdxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/INTL at: <http://image.llnl.gov>
Series: IRAK Plate: 125 Row: m Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21311987.

FEATURES	Location/Qualifiers
source	1..2087

```

/note="Synonyms: 4930463F05RLX, 2810017F12RLX
/db xref="LocusID:74026"
/db xref="MGI:1921276"
15. .1406

```

```

/product="4121402D02Rik_protein"
/protein_id="AAH55715.1"
/db_xref="LOC:3158902"
/db_xref="GeneID:74026"
/db_xref="WGI:1921276"
/translation="MTMR5AVFKAAAPAGCGNPEQRLDYERAAALGSPEDESGAABAH
FLRHRRLKEPPELSSQGSSPSPAGCGGGGRLLPAGAAPQOEWSWGASVPL
LPCEPATKQAGIGGEVVAAGAGSPRKTAQAVLPITQGISVVAAPETPMADKDKG
AAPPAATAPDAGPPEPLPPEPPPLATAGTTLAABEGWKSIRKPLGGCGSGAS
SOAAKQIILLDLBOQOOLQAAKEIEBLSKSRDLLIEEMERMOVLKRD
NEKERHLLGGVDEPEREAELSEKILKEBLOPELCSQALLPSKPSGSGSKHKKR
TPGONTERTKTPVKKLABEPSKVTKTKPKSHPIKEEPGSGISBYCKRBLRSQETPEK
RSSVDVPEPLRSTPQKGPSTHPKEKAFSEMEDLPYLSITTEYLCRMHQPPSPPLABE
SSPEKKEETVASKA"

```

Query Match 2.78; Score 116; DB 10; Length 2087;

[illegible]

RESULT 41	LOCUS	DEFINITION
BC043039	2095 bp	linear
BC043039	2095 bp	linear
Mus musculus RIKEN	CDNA	ROD 21-OCT-2003
	4121402D02 gene,	mRNA (CDNA clone MGC:579934

IMAGE:5701204), complete cds.
ACCESSION BC043039
VERSION BC043039.1 GI:27696670
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 2095)
Strausberg,R.L., Feligold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diachenko,L., Marusina,K., Farmer,M.F., Casavant,T.L., Scaplehorn,W., Soares,M.B., Bonaldo,M.A., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carinci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahney,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2388257
TITLE 2 (bases 1 to 2095)
Strausberg,R.
REFERENCE Direct Submission
JOURNAL Submitted (09-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr. Thomas L. Casavant.
Web site: <http://genome.uiowa.edu>
Contact: bento-soares@uiowa.edu
Bonaldo,M.F., Akabogu,I., Bait,T., Bait,J., Crouch,K., Davis,A., Fishler,K., Keppel,C., Kucada,T., Lebeck,M., Melo,A., Schaefer,K., Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J., Casavant,T., Soares,M.B.
FEATURES
source
1..2095
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:57934 IMAGE:5701204"
/tissue_type="Brain, mouse 15.5 dpc"
/clone_lib="NIH_BMAP_EVO"
/lab_host="DH10B"
/note="Vector: pYX-ASC"
1..2095
/gene="4121402D02Rik"
/note="synonyms: 4930463F05Rik, 281001712Rik"

CDs
/db_xref="LOCUSID:74026"
/db_xref="MGI:1921276"
41..1432
/gene="4121402D02Rik"
/codon_start=1
/product="4121402D02Rik protein"
/protein_id="AAH43039.1"
/db_xref="GI:27696671"
/db_xref="LOCUSID:74026"
/db_xref="MGI:1921276"
/translation="MTMRSAVPKAAAPAGNPNORLDYERAAALGPPDESGAEN
FIPRRRLKEPPPLASQSGSGGKGGLLPGAAPQOEGSGSV
LCPSPATQKQAGGSPRYQVLPQTSIVYAAKEPVPAGDGG
AAPPATADDPAGPPPLPSPPPPLAPTATGTLASBERMWSIRSPICGGGSGAS
SQDACQQLQLDLEQQQQLQKEIELESDTLARIMERMQLVRD
NFKERKTLQGEPEPERAEISEKIKERQPELCEQALPSKPSGSGKHRR
TNGKNTKTPYKALPEPSKYTKTPKPSPIKEPSCGISFVCKRELRSDTPPKP
RSSVDPPLSTPQKPSHPREKAFSSMEDLPYLSTTEMYLCRWHPPPPLPLRE
SSPKKEETVASSXA"

ORIGIN

Query Match 2.7%; Score 116; DB 10; Length 2095;
Best Local Similarity 100.0%; Pred. No. 2e-56;
Matches 116; Conservative 0; Mismatches 0; Gaps 0;

QY 1338 GAAGATTGCGCGTACCTTCCACCAAGAAATGTTGTCGTGGACCGCCTCC 1397
DB 1307 GAAGATTGCGCGTACCTTCCACCAAGAAATGTTGTCGTGGACCGCCTCC 1366
1398 CCATCACCGTTACCTTCCACCAAGAAATGTTGTCGTGGACCGCCTCC 1453
DB 1367 CCATCACCGTTACCTTCCACCAAGAAATGTTGTCGTGGACCGCCTCC 1422

RESULT 42
LOCUS AX401049 615 bp DNA linear PAT 06-JUN-2002
DEFINITION Sequence 725 from Patent WO0210453.
ACCESSION AX401049
VERSION AX401049.1 GI:21337229
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1
AUTHORS Mendrick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and
Elaasroff,M.R.
TITLE Molecular toxicology modeling
JOURNAL Patent: WO 0210453-A 725 07-FEB-2002;
Gene Logic, Inc. (US)
FEATURES
source
1..615
/organism="Rattus norvegicus"
/mol_type="unassigned DNA"
/db_xref="taxon:10116"
/note="EMBL/Genbank Accession No. A1102093"

ORIGIN

Query Match 2.5%; Score 110; DB 6; Length 615;
Best Local Similarity 100.0%; Pred. No. 6.1e-53;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 GGCTGACGCGCGCGGCAAGTATCAGGCGGTCTGCCATTGACAGCGGCTCTCGTG 506
DB 339 GGCTGACGCGCGCGGCAAGTATCAGGCGGTCTGCCATTGACAGCGGCTCTCGTG 398
QY 507 GCGCGCGGCAAGAGCTACGCGCTGCGGTGGGACAGAGGTTGGCGCGC 556
DB 399 GCGCGCGGCAAGAGCTACGCGCTGCGGTGGGACAGAGGTTGGCGCGC 448

RESULT 43
AX913485 317 bp DNA linear PAT 18-DEC-2003
LOCUS
DEFINITION Sequence 29348 from Patent EP1033401.
ACCESSION AX913485
VERSION AX913485.1 GI:40207274
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Dumas Milne Edwards,J.B., Duclaire,A. and Giordano,J.Y.
AUTHORS
TITLE Expressed sequence tags and encoded human proteins
JOURNAL Patent: EP 1033401-A 29348 06-SEP-2000;
Genet (FR)
FEATURES
Location/Qualifiers
1..317
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 2.4%; Score 106; DB 6; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.3e-50;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3569 TAGAAAACTGGGCGAGATTTCTTGTCTCCATCATTTTAATGCGGAGGCTTCAG 3628
Db 212 TAGAAAACTGGGCGAGATTTCTTGTCTCCATCATTTTAATGCGGAGGCTTCAG 271
Qy 3629 TTTTCTTACTCTTACCTATGATGATATTTCTTCGTAAGTCGCCAA 3674
Db 272 TTTTCTTACTCTTACCTATGATGATATTTCTTCGTAAGTCGCCAA 317
RESULT 44
BD049018 317 bp DNA linear PAT 27-AUG-2002
LOCUS
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD049018
VERSION BD049018.1 GI:22590760
KEYWORDS JP 2001269182-A/25264.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 317)
AUTHORS Edwards,J.B.D.M., Duclaire,E. and Jordan,J.Y.
TITLE Sequence tag and encoded human protein
JOURNAL Patent: JP 2001269182-A 25264 02-OCT-2001;
GENET
OS Homo sapiens (human)
PN JP 2001269182-A/25264
PD 02-OCT-2001
PR 24-FEB-2000 JP 2000118773
PR 26-FEB-1999 US 60/122487
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES
PT JORDAN
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10,
PC C12P21/02, C12P21/08, C12Q1/68//G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC
FH Key Location/Qualifiers
1..317
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 2.4%; Score 106; DB 6; Length 317;

Best Local Similarity 100.0%; Pred. No. 1.3e-50;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3569 TAGAAAACTGGGCGAGATTTCTTGTCTCCATCATTTTAATGCGGAGGCTTCAG 3628
Db 212 TAGAAAACTGGGCGAGATTTCTTGTCTCCATCATTTTAATGCGGAGGCTTCAG 271
Qy 3629 TTTTCTTACTCTTACCTATGATGATATTTCTTCGTAAGTCGCCAA 3674
Db 272 TTTTCTTACTCTTACCTATGATGATATTTCTTCGTAAGTCGCCAA 317
RESULT 45
BV207730 201 bp DNA linear STS 10-JUN-2004
LOCUS
DEFINITION sqm224404 Human DNA (Sequenc) Homo sapiens STS genomic, sequence
tagged site.
ACCESSION BV207730
VERSION BV207730.1 GI:48178296
KEYWORDS STS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 201)
AUTHORS Nelson,R.M., Marnellos,G., Kammerer,S., Hoyal,C.R., Shi,M.M.,
Cantor,C.R. and Braun,A.
TITLE Large-Scale Validation of Single Nucleotide Polymorphisms in Gene
Regions
JOURNAL Genome Res. (2004) In press
COMMENT
Contact: Andreas Braun
Pharmaceuticals division
Sequenom, Inc.
3595 John Hopkins Court, San Diego, CA 92121, USA
Tel: 18582029018
Fax: 18582029020
Email: abraun@sequenom.com
Primer A: No primer sequence submitted
Primer B: No primer sequence submitted
STS size: 201.
FEATURES
Location/Qualifiers
1..201
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="Human DNA (Sequenc)"
ORIGIN
STS
Query Match 1.9%; Score 85; DB 11; Length 201;
Best Local Similarity 100.0%; Pred. No. 3.7e-38;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4258 GATACATTTCTGATGTAATAATGTTGCTGTGAATATCTTAAAGAGTCAAT 4317
Db 1 GATACATTTCTGATGTAATAATGTTGCTGTGAATATCTTAAAGAGTCAAT 60
Qy 4318 GTAAATTAACCTATGTCGTGTTAA 4342
Db 61 GTAAATTAACCTATGTCGTGTTAA 85
RESULT 46
CQ552685 60 bp DNA linear PAT 30-JAN-2004
LOCUS
DEFINITION Sequence 22320 from Patent WO0210449.
ACCESSION CQ552685
VERSION CQ552685.1 GI:41519112
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE          Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
AUTHORS            Shoshan,A., Maeserman,A., Mintz,E., Mintz,L. and Faigler,S.
TITLE              Oligonucleotide library for detecting rna transcripts and splice
JOURNAL            variants that populate a transcritpome
                   Patent: WO 0210449-A 22320 07-FEB-2002;
FEATURES            CompuGen Inc. (US)
                   Location/Qualifiers
                   1..60
                   /organism="Homo sapiens"
                   /mol_type="unassigned DNA"
                   /db_xref="taxon:9606"
ORIGIN
Query Match       1.4%; Score 60; DB 6; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.2e-23;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY                1201 CTTGTGTAACGTGATGAGAGCCAGAAACCCCGGCTTCAGTGG 1260
Db                1 CTTGTGTAACGTGATGAGAGCCAGAAACCCCGGCTTCAGTGG 60

RESULT 47
AC107643/C        208620 bp      DNA      linear      HTG 22-MAR-2004
LOCUS              Mus musculus chromosome 14 clone RP23-54L14 map 14, *** SEQUENCING
DEFINITION         IN PROGRESS ***, 7 ordered pieces.
AC107643
AC107643.4 GI:45599009
VERSION            HTG; HTGS PHASE2; HTGS_FULFILLTOP; HTGS_ACTIVEFIN.
KEYWORDS           Mus musculus (house mouse)
SOURCE             Mus musculus
ORGANISM           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE          1 (bases 1 to 208620)
AUTHORS            Birren,B., Nusbaum,C. and Lander,E.
TITLE              Mus musculus chromosome 14, clone RP23-54L14
JOURNAL            Unpublished
AUTHORS            2 (bases 1 to 208620)
                   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
                   Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B.,
                   Brown,A., Camarata,J., Campopiano,A., Chang,J., Charazot,B.,
                   Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
                   Cooke,P., Dearlailan,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
                   Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
                   Garg,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
                   Hagoes,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
                   Kamae,A., Karakas,A., Kells,C., Lacroque,K., Lamazares,R.,
                   Landers,I., Lenoczky,J., Levine,R., Liu,G., Maclean,C.,
                   Macdonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M.,
                   McEwan,P., McKernan,K., Meldrum,J., Menues,L., Mihova,T.,
                   Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
                   Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
                   Petersen,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
                   Retta,R., Ribbeck,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
                   Rossetti,M., Roy,A., Santos,R., Schauer,N., Schjorback,R., Seaman,S.,
                   Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
                   Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
                   Thomas,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
                   Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
                   Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
JOURNAL            Direct Submission
REFERENCE          Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
                   Research, 320 Charles Street, Cambridge, MA 02141, USA
TITLE              3 (bases 1 to 208620)
JOURNAL            Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
                   Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
                   Bogunavsky,L., Bouckgalter,B., Camarata,J., Chang,J., Choepel,Y.,
                   Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
                   Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fero,S.,
                   Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
                   Graham,L., Grand-Pierre,N., Hagoes,B., Hagopian,D., Hagoes,B.,

```

```

Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamae,A., Karakas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Liu,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Menues,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramaamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schjorback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
JOURNAL            Direct Submission
COMMENT            Submitted (22-MAR-2004) Whitehead Institute/MIT Center for Genome
                   Research, 320 Charles Street, Cambridge, MA 02141, USA
                   On Mar 22, 2004 this sequence version replaced gi:29171454.
                   All repeats were identified using RepeatMasker:
                   Smit, A.P.A. & Green, P. (1996-1997)
                   http://ftp.genome.washington.edu/RM/RepeatMasker.html
                   ----- Genome Center
                   Center: Whitehead Institute/ MIT Center for Genome Research
                   Center code: WIBR
                   Web site: http://www-seq.wi.mit.edu
                   Contact: sequence_submissions@genome.wi.mit.edu
                   ----- Project Information
                   Center project name: L14249
                   Center clone name: 54_L_14
                   -----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 96362: contig of 96362 bp in length
* 96363 96462: gap of 100 bp
* 96463 99202: contig of 2740 bp in length
* 99203 99302: gap of 100 bp
* 99303 155557: contig of 56255 bp in length
* 155558 156557: gap of 100 bp
* 156558 165735: contig of 10078 bp in length
* 165736 165835: gap of 100 bp
* 165836 168011: contig of 2176 bp in length
* 168012 168111: gap of 100 bp
* 168112 173215: contig of 5104 bp in length
* 173216 173315: gap of 100 bp
* 173316 208620: contig of 35305 bp in length.
FEATURES            Location/Qualifiers
                   1..208620
                   /organism="Mus musculus"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:10090"
                   /chromosome="14"
                   /map="14"
                   /clone="RP23-54L14"
                   /clone_1b="RP23-54L14"
ORIGIN              1.3%; Score 55; DB 2; Length 208620;
Query Match        Best Local Similarity 100.0%; Pred. No. 8.9e-20;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY                1098 AAGACTCTGTAAAGCTGCGCTCTGATTTTCAAAAGTCAAAACAAACCTC 1152
Db                195356 AAGACTCTGTAAAGCTGCGCTCTGATTTTCAAAAGTCAAAACAAACCTC 195302

RESULT 48
AX400340/c

```

LOCUS AX400340 590 bp DNA linear PAT 06-JUN-2002
 DEFINITION Sequence 16 from Patent WO0210453.
 AX400340
 VERSION AX400340.1 GI:21336520
 KEYWORDS
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 AUTHORS Mendick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and
 Elashoff,M.R.
 TITLE Molecular toxicology modeling
 JOURNAL Patent: WO 0210453-A 16 07-FEB-2002;
 Gene Logic, Inc. (US)
 FEATURES
 source
 1. 590
 /organism="Rattus norvegicus"
 /mol_type="unassigned DNA"
 /db_xref="taxon:10116"
 /note="EMBL/Genbank Accession No. AA799599"

ORIGIN
 Query Match 1.2%; Score 51; DB 6; Length 590;
 Best Local Similarity 100.0%; Pred. No. 7.5e-18;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4286 TGTCTTGTAAATATCTTAATAAGAGTTCAATTGTAATAAATCTATTGTGCGC 4336
 55 TGTCTTGTAAATATCTTAATAAGAGTTCAATTGTAATAAATCTATTGTGCGC 5

RESULT 49
 LOCUS BC043668 615 bp mRNA linear ROD 13-FEB-2004
 DEFINITION Mus musculus cDNA clone IMAGE:5029752, partial cds.
 BC043668
 VERSION BC043668.1 GI:27693990
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,J., Shenmen,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haieh,F.,
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Ushed,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullah,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J., Helton,E., Kettman,M., Madan,A., Cordero,R.A.,
 Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butterfield,Y.S., Krzywinski,M.T., Skalska,U., Smalins,D.E.,
 Schnerch,A., Schein,U.B., Jones,S.J. and Marra,M.A.
 TITLE Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 PUBMED 12477932
 REFERENCE
 2 (bases 1 to 615)
 Strausberg,R.
 DIRECT SUBMISSION
 Submitted (10-JUN-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Lohar Hemighausen Ph.D., Robin Humphreys
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@cgsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
 Andy Chan, Steve S. Chand, William Chow, Allison Cloutier, Ruth
 Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
 Kim Macdonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
 Teika Olson, Diana Palmquist, Anca Petrescu, Anna Lisa Prabh,
 Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
 Duane Smalins, Jeff Scott, Miranda Tsai, George Yang, Jacqui
 Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRK Plate: 86 Row: C Column: 22.

FEATURES
 source
 1. 615
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="mix FVB/N, C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:5029752"
 /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
 old, gross tissue."
 /clone_idb="NCI CGAP_Mam5"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"

ORIGIN
 Query Match 1.2%; Score 51; DB 10; Length 615;
 Best Local Similarity 100.0%; Pred. No. 7.6e-18;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4286 TGTCTTGTAAATATCTTAATAAGAGTTCAATTGTAATAAATCTATTGTGCGC 4336
 106 TGTCTTGTAAATATCTTAATAAGAGTTCAATTGTAATAAATCTATTGTGCGC 156

RESULT 50
 LOCUS AX899390 50 bp DNA linear PAT 18-DEC-2003
 DEFINITION Sequence 15253 from Patent EP1033401.
 AX899390
 VERSION AX899390.1 GI:40054303
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Dumas Milne Edwards,J.B., Duclet,A. and Giordano,J.Y.
 TITLE Expressed sequence tags and encoded human proteins
 JOURNAL Patent: EP 1033401-A 15253 06-SEP-2000;
 Genset (FR)
 FEATURES
 source
 1. 50
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 1.1%; Score 47; DB 6; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.2e-15;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 AAACACTCACCGAGCTGTGAATAATAGCTGTGTGCAAGAACCC 1940

DB 1 AAACACCTCACCGACGTGTAGGAATAAGCTGTCTGCGCAAGAACCC 47

RESULT 51
BD034923
LOCUS
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BD034923 50 bp DNA linear PAT 27-AUG-2002
Sequence tag and encoded human protein.
BD034923.1 GI:22576665
JP 2001269182-A/11169.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 50)
Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
Sequence tag and encoded human protein
Patent: JP 2001269182-A 11169 02-OCT-2001;
GENSET

COMMENT
OS Homo sapiens (human)
PN JP 2001269182-A/11169
PD 02-OCT-2001
PF 24-FEB-2000 JP 2000118773
PR 26-FEB-1999 US 60/122487
PI JEAN BAPTISTE DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
FI JORDAN
PC C12N5/10,
PC C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
PC C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
G06F15/40
CC
FH Key Location/Qualifiers.
CC 1.50
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
LOCATION/Qualifiers
1.50
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 1.1%; Score 47; DB 6; Length 50;
Best local Similarity 100.0%; Pred. No. 1.2e-15;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 AAACACCTCACCGACGTGTAGGAATAAGCTGTCTGCGCAAGAACCC 1940
1 AAACACCTCACCGACGTGTAGGAATAAGCTGTCTGCGCAAGAACCC 47

Db
RESULT 52
AL928900/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL928900 312459 bp DNA linear HTG 24-SEP-2003
Danio rerio clone CH211-138M15, *** SEQUENCING IN PROGRESS ***, 45
unordered pieces.
AL928900 GI:33411120
HTG; HTGS_PHASE1.
Danio rerio (zebrafish)
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 312459)
Plumb,B.
Direct Submission
Submitted (24-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 1, 2003 this sequence version replaced gi:25136732.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Contact: zfish-help@sanger.ac.uk
Center project name: zC138M15
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 295273 bases at least Q40
Consensus quality: 300876 bases at least Q30
Consensus quality: 304815 bases at least Q20
Insert size: 308059; sum-of-contigs
Insert size: 183497; 4.5% error; agarose-fp
Quality coverage: 6.17x in Q20 bases; sum-of-contigs Quality
coverage: 13.54x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 18133: contig of 18133 bp in length
18134 18233: gap of 100 bp
18234 25392: contig of 7159 bp in length
25393 25492: gap of 100 bp
25493 27992: contig of 2500 bp in length
27993 28092: gap of 100 bp
28093 30488: contig of 2396 bp in length
30489 30588: gap of 100 bp
30589 35150: contig of 4562 bp in length
35151 35250: gap of 100 bp
35251 37593: contig of 2343 bp in length
37594 37693: gap of 100 bp
37694 39774: contig of 2081 bp in length
39775 39874: gap of 100 bp
39875 42550: contig of 2676 bp in length
42551 42650: gap of 100 bp
42651 45058: contig of 2408 bp in length
45059 45158: gap of 100 bp
45159 51510: contig of 6352 bp in length
51511 51610: gap of 100 bp
51611 56050: contig of 4440 bp in length
56051 56150: gap of 100 bp
56151 60678: contig of 4528 bp in length
60679 60778: gap of 100 bp
60779 65716: contig of 4938 bp in length
65717 65817: gap of 100 bp
65817 71354: contig of 5538 bp in length
71355 71454: gap of 100 bp
71455 74067: contig of 2613 bp in length
74068 74167: gap of 100 bp
74168 76704: contig of 2537 bp in length
76705 76804: gap of 100 bp
76805 79615: contig of 2811 bp in length
79616 79715: gap of 100 bp
79715 83397: contig of 3682 bp in length
83398 83497: gap of 100 bp
83498 86169: contig of 2672 bp in length
86169 86269: gap of 100 bp
86270 90191: contig of 3922 bp in length
90192 90291: gap of 100 bp
90292 92810: contig of 2519 bp in length
92811 92910: gap of 100 bp
92911 95265: contig of 2355 bp in length
95266 95365: gap of 100 bp
95366 97646: contig of 2281 bp in length
97647 97746: gap of 100 bp
97747 100395: contig of 2649 bp in length
100396 100495: gap of 100 bp
100496 103820: contig of 3325 bp in length
103821 103920: gap of 100 bp
103921 110883: contig of 6963 bp in length

```
* 110884 110983: gap of 100 bp
* 110984 113098: contig of 2115 bp in length
* 113098 113198: gap of 100 bp
* 113199 121384: contig of 8186 bp in length
* 121385 121484: gap of 100 bp
* 121485 125791: contig of 4307 bp in length
* 125792 125891: gap of 100 bp
* 125892 128931: contig of 3040 bp in length
* 129032 129031: gap of 100 bp
* 129032 135415: contig of 6384 bp in length
* 135416 135515: gap of 100 bp
* 135516 142402: contig of 6887 bp in length
* 142403 142502: gap of 100 bp
* 142503 144608: contig of 2106 bp in length
* 144609 144708: gap of 100 bp
* 144709 147437: contig of 2729 bp in length
* 147438 147537: gap of 100 bp
* 147538 150315: contig of 2778 bp in length
* 150316 150415: gap of 100 bp
* 150416 153357: contig of 2942 bp in length
* 153358 153457: gap of 100 bp
* 153458 158644: contig of 5187 bp in length
* 158645 158744: gap of 100 bp
* 158745 163851: contig of 5107 bp in length
* 163852 163951: gap of 100 bp
* 163952 166599: contig of 2648 bp in length
* 166600 166699: gap of 100 bp
* 166700 169492: contig of 2793 bp in length
* 169493 169592: gap of 100 bp
* 169593 207022: contig of 37430 bp in length
* 207023 207122: gap of 100 bp
* 207123 230362: contig of 23240 bp in length
* 230363 230462: gap of 100 bp
* 230463 272343: contig of 41861 bp in length
* 272344 272443: gap of 100 bp
* 272444 293830: contig of 21387 bp in length
* 293831 293930: gap of 100 bp
* 293931 312459: contig of 18529 bp in length.
```

```
FEATURES
Source
1. .312459
Location/Qualifiers
```

```
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-138M15"
/clone_lib="CHOR1-211"
1. .18133
/note="assembly_fragment:00013
fragment_chain:1"
18234. .25392
/note="assembly_fragment:00376
fragment_chain:1"
25493. .27992
/note="assembly_fragment:00981
fragment_chain:1"
28093. .30488
/note="assembly_fragment:00572
fragment_chain:2"
30589. .35150
/note="assembly_fragment:04315
fragment_chain:2"
35251. .37593
/note="assembly_fragment:00766
fragment_chain:3"
37694. .39774
/note="assembly_fragment:04156
fragment_chain:3"
39875. .42550
/note="assembly_fragment:01094
fragment_chain:4"
42651. .45058
/note="assembly_fragment:01244
fragment_chain:4"
45159. .51510
```

```
misc_feature /note="assembly_fragment:00022"
51611. .56050
/note="assembly_fragment:00412"
misc_feature /note="assembly_fragment:00445"
56151. .60678
60779. .65716
/note="assembly_fragment:00502"
misc_feature /note="assembly_fragment:00554"
65817. .71354
71455. .74067
/note="assembly_fragment:00566"
misc_feature /note="assembly_fragment:00576"
74168. .76704
76805. .79615
/note="assembly_fragment:00885"
misc_feature /note="assembly_fragment:00997"
79716. .83397
83498. .86169
/note="assembly_fragment:01087"
misc_feature /note="assembly_fragment:01134"
86270. .90191
90292. .92810
/note="assembly_fragment:01145"
misc_feature /note="assembly_fragment:01173"
92911. .95265
95366. .97646
/note="assembly_fragment:01199"
misc_feature /note="assembly_fragment:01218"
97747. .100395
100496. .103820
/note="assembly_fragment:01608"
misc_feature /note="assembly_fragment:01608"
103921. .110883
```

```
Query Match 0.9%; Score 40; DB 2; Length 312459;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Dy 4333 TGCGCTGTAAAAA... 4372
Db 98146 TGCGCTGTAAAAA... 98107

RESULT 53
AC140799 162329 bp DNA linear ROD 29-JUN-2004
DEFINITION Mus musculus chromosome 15, clone RP24-147N23, complete sequence.
AC140799
VERSION AC140799.6 GI:49406059
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 162329)
2 (bases 1 to 162329)
JOURNAL Mus musculus chromosome 15, clone RP24-147N23
TITLE Unpublished
AUTHORS Birren,B., Nusbaum,C., Lander,B., Aboueleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Baerlien,V., Bloom,T.,
Boguslavsky,L., Boukhalter,B., Canarata,J., Chang,J., Choepel,Y.,
Collimore,A., Cook,A., Cooke,P., Corum,B., D'Arcangelo,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzderald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karakas,A., Kells,C., Lander,B., Levene,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
```

Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 162329)
Birten,B., Nusbaum,C., Lander,E., Aboueljelil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Bouglavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., DeArliano,K., Diaz,U.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,U., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhang,P., Pierre,N., Rachupka,A., Ramsamy,U., Raymond,C., Retia,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,U., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-APR-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 162329)
Birten,B., Nusbaum,C., Lander,E., Aboueljelil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Bouglavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Corum,B., DeArliano,K., Diaz,U.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,U., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhang,P., Pierre,N., Rachupka,A., Ramsamy,U., Raymond,C., Retia,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,U., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (29-JUN-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 29, 2004 this sequence version replaced gt:4635993.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@road.mit.edu
----- Project Information
Center project name: L24402
Center clone name: 147_N_23

Location/Qualifiers
1..162329

misc_feature
1..4
/note="clone boundary
clone_end:SP6
site:MboI
651..679
/rpt_family="AT_rich"
complement(1253..1363)
/rpt_family="BCL_MM"
1872..2596
/rpt_family="Lx9"
2599..2679
/rpt_family="(CA)n"
3153..3202
/rpt_family="(TC)n"
3204..3257
/rpt_family="(TC)n"
3309..3342
/rpt_family="(TTTG)n"
3348..3371
/rpt_family="(TTTG)n"
complement(5329..5479)
/rpt_family="MTE"
7524..7564
/rpt_family="MIR3"
complement(9579..9815)
/rpt_family="RMER16"
9883..9918
/rpt_family="(T)n"
complement(9919..10060)
/rpt_family="B1_MM"
complement(10055..10176)
/rpt_family="RMER16"
10159..10266
/rpt_family="RSINE1"
10270..10349
/rpt_family="(CA)n"
10872..11011
/rpt_family="(CA)n"
12031..12061
/rpt_family="(CAA)n"
12576..12652
/rpt_family="(GGA)n"
13239..13399
/rpt_family="B2_Mm1"
13433..13540
/rpt_family="B2_Mm2"
13694..13721
/rpt_family="AT_rich"
14544..14594
/rpt_family="(TC)n"
14617..14678
/rpt_family="CT_rich"
14778..15040
/rpt_family="MLT2B3"
complement(15479..15550)
/rpt_family="MTD"
16026..16417
/rpt_family="RMERS5"
17328..17406
/rpt_family="(GA)n"
17447..17586
/rpt_family="(TCTCC)n"
complement(17487..17529)
/note="single clone coverage"
complement(19544..19757)
/rpt_family="LIMed"


```

repeat_region complement(20049..20404)
                /rpt_family="ORR1C"
repeat_region complement(20628..20902)
                /rpt_family="RMER3"
repeat_region 21215..21246
                /rpt_family="AT_rich"
repeat_region 21833..22076
                /rpt_family="MTE"
repeat_region 22867..23199
                /rpt_family="ORR1B-int"
repeat_region complement(23207..23388)
                /rpt_family="B3A"
repeat_region 23396..23567
                /rpt_family="ORR1B-int"
repeat_region 23631..23825
                /rpt_family="MLT2B3"
repeat_region 24404..24446
                /rpt_family="CA)n"
repeat_region complement(24447..24576)
                /rpt_family="LMED"
repeat_region 24641..24670
                /rpt_family="CAA)n"
repeat_region complement(24930..25253)
                /rpt_family="ORR1A3"
repeat_region 25265..25321

Query Match      0.9% Score 39; DB 10; Length 162329;
Best Local Similarity 100.0%; Pred. No. 2.8e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4291 TGTAAATATCTTATTAAGAGCTTCATTGTAAATTAACCTA 4329
Db 18603 TGTAAATATCTTATTAAGAGCTTCATTGTAAATTAACCTA 18641

RESULT 54
AC102314 198220 bp DNA linear HTG 21-AUG-2002
LOCUS Mus musculus clone RP24-439D9, WORKING DRAFT SEQUENCE, 36 unordered
DEFINITION pieces.
AC102314 AC102314.2 GI:22381118
VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 198220)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-439D9
Unpublished
2 (bases 1 to 198220)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Charazot,B.,
Chopel,Y., Cologelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArliano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,K., Heath,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Labrecque,K.,
Lamaras,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Riebach,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,N., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 198220)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouhgalter,B.,
Camarata,J., Chang,J., Charazot,B., Chopel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArliano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fato,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenka,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,N., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Strange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17061400.
All repeats were identified using RepeatMasker:
Smith,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L18338
Center Clone name: 439 D.9
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 187584 bases at least Q40
Consensus quality: 191843 bases at least Q30
Consensus quality: 193381 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 194720; sum-of-ctab
Quality coverage: 6.3 in Q20 bases; agarose-fp
Quality coverage: 6.1 in Q20 bases; sum-of-ctab
-----
NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1117: contig of 1117 bp in length
1118 1217: gap of 100 bp
1218 2402: contig of 1185 bp in length
2403 2502: gap of 100 bp
2503 3868: contig of 1366 bp in length
3869 3968: gap of 100 bp
3969 5406: contig of 1438 bp in length
5407 5507: gap of 100 bp
5508 6534: contig of 1028 bp in length
6535 9062: gap of 100 bp
9063 9162: contig of 2428 bp in length
9163 10579: gap of 100 bp
10580 10679: contig of 1417 bp in length
10680 13095: contig of 2416 bp in length

```

* 13096 13195: gap of 100 bp
* 13196 15395: contig of 2200 bp in length
* 15396 15495: gap of 100 bp
* 15496 17871: contig of 2376 bp in length
* 17872 20551: contig of 2580 bp in length
* 20552 22778: contig of 2127 bp in length
* 22779 22878: gap of 100 bp
* 22879 25337: contig of 2443 bp in length
* 25338 25427: gap of 100 bp
* 25428 27465: contig of 2038 bp in length
* 27466 27565: gap of 100 bp
* 27566 29349: contig of 1784 bp in length
* 29350 29449: gap of 100 bp
* 29450 32095: contig of 2646 bp in length
* 32096 32195: gap of 100 bp
* 32196 34776: contig of 2581 bp in length
* 34777 34876: gap of 100 bp
* 34877 37945: contig of 3069 bp in length
* 37946 38045: gap of 100 bp
* 38046 42221: contig of 4176 bp in length
* 42222 42321: gap of 100 bp
* 42322 46116: contig of 3795 bp in length
* 46117 50462: gap of 100 bp
* 50463 50562: gap of 100 bp
* 50563 54508: contig of 3946 bp in length
* 54509 54608: gap of 100 bp
* 54609 59386: contig of 4776 bp in length
* 59387 59486: gap of 100 bp
* 59487 65412: contig of 5926 bp in length
* 65413 65512: gap of 100 bp
* 65513 72602: contig of 7090 bp in length
* 72603 78492: contig of 5790 bp in length
* 78493 78592: gap of 100 bp
* 78593 84244: contig of 5652 bp in length
* 84245 84344: gap of 100 bp
* 84345 92188: contig of 7844 bp in length
* 92189 92288: gap of 100 bp
* 92289 99017: contig of 6729 bp in length
* 99018 99117: gap of 100 bp
* 99118 111214: contig of 12097 bp in length
* 111215 111314: gap of 100 bp
* 111315 121265: contig of 9951 bp in length
* 121266 121365: gap of 100 bp
* 121366 134090: contig of 12725 bp in length
* 134091 134190: gap of 100 bp
* 134191 144836: contig of 10646 bp in length
* 144837 144936: gap of 100 bp
* 144937 158280: contig of 13344 bp in length
* 158281 158380: gap of 100 bp
* 158381 176170: contig of 17790 bp in length
* 176171 176270: gap of 100 bp
* 176271 198220: contig of 21950 bp in length.
* 198220 Location/Qualifiers
1. 198220
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_lib="RP24-439D9"
/clone_1ib="RP24-439D9"
1. 1117
/note="assembly_fragment"
1218. 2402
/note="assembly_fragment"
2503. 3868
/note="assembly_fragment"
3969. 5406
/note="assembly_fragment"
5507. 6534
/note="assembly_fragment"

misc_feature 6635..9062 /note="assembly_fragment"
misc_feature 9163..10579 /note="assembly_fragment"
misc_feature 10680..13095 /note="assembly_fragment"
misc_feature 13196..15395 /note="assembly_fragment"
misc_feature 15496..17871 /note="assembly_fragment"
misc_feature 17972..20551 /note="assembly_fragment"
misc_feature 20652..22778 /note="assembly_fragment"
misc_feature 22879..25337 /note="assembly_fragment"
misc_feature 25428..27465 /note="assembly_fragment"
misc_feature 27466..29349 /note="assembly_fragment"
misc_feature 29350..29449 /note="assembly_fragment"
misc_feature 29450..32095 /note="assembly_fragment"
misc_feature 32096..32195 /note="assembly_fragment"
misc_feature 32196..34776 /note="assembly_fragment"

Query Match 0.9%; Score 39; DB 2; Length 198220;
Best Local Similarity 100.0%; Pred.No. 2.9e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4334 GGCTGTTAAAAA... 4372
Db 7792 GGCTGTTAAAAA... 7830

RESULT 55
AC133452/c 324084 bp DNA linear HTG 23-APR-2003
LOCUS Mus musculus clone RP24-14814, WORKING DRAFT SEQUENCE, 21 unordered
DEFINITION
AC133452
AC133452
VERSION
HTG: HTGS PHASE1; HTGS DRAFT.
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCES
AUTHORS
2 (bases 1 to 324084)
Birren,B., Nubbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horcun,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneses,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupbach,R., Seaman,C., Severy,P.,
Smith,C., Spencer,B., Strange-Thomann,N., Stojanovic,N., Talamas,J.,
Testafaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCES
AUTHORS
Submitted (12-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 324084)
Birren,B., Nubbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,

TITLE
JOURNAL
COMMENT

Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kemat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, D., Menus, L., Mihova, T., Miska, V., Murphy, J., Naylor, J., Nguyen, C., Nicol, R., Nordu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhara, P., Pierre, N., Raghupathi, A., Ramasamy, U., Raymond, C., Retter, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strubbe, M., Talamas, J., Testa, S., Theodore, J., Tophan, K., Travers, M., Vassiliou, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.

Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 17, 2003 this sequence version replaced gi:28416226.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: sequence_submission@genome.wi.mit.edu

Project Information

Center project name: L26549

Center clone name: 148_1_4

NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1011: contig of 1011 bp in length
* 1012 1111: gap of 100 bp
* 1112 2019: contig of 908 bp in length
* 2020 2119: gap of 100 bp
* 2120 2920: contig of 801 bp in length
* 2921 3021: gap of 100 bp
* 3021 4680: contig of 1660 bp in length
* 4681 4781: gap of 100 bp
* 4781 6502: contig of 1722 bp in length
* 6503 6602: gap of 100 bp
* 6603 8777: contig of 2175 bp in length
* 8778 8877: gap of 100 bp
* 8878 11203: contig of 2326 bp in length
* 11204 11303: gap of 100 bp
* 11304 14001: contig of 2698 bp in length
* 14002 14101: gap of 100 bp
* 14102 16893: contig of 2792 bp in length
* 16894 16993: gap of 100 bp
* 16994 21284: contig of 4291 bp in length
* 21285 21385: gap of 100 bp
* 21385 25083: contig of 3699 bp in length
* 25084 302183: gap of 100 bp
* 302184 33023: contig of 7840 bp in length
* 33024 33123: gap of 100 bp
* 33124 43063: contig of 9940 bp in length
* 43064 43163: gap of 100 bp
* 43164 56342: contig of 13179 bp in length
* 56343 56442: gap of 100 bp
* 56443 75509: contig of 19067 bp in length
* 75510 75609: gap of 100 bp
* 75610 90469: contig of 14860 bp in length
* 90470 90569: gap of 100 bp
* 90570 109549: contig of 18980 bp in length
* 109550 109649: gap of 100 bp
* 109650 160746: contig of 51097 bp in length

```

```

* 160747 160846: gap of 100 bp
* 160847 227679: contig of 66833 bp in length
* 227680 227779: gap of 100 bp
* 227780 285533: contig of 57754 bp in length
* 285534 285633: gap of 100 bp
* 285634 324084: contig of 38451 bp in length.
Location/Qualifiers
1..324084
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-14814"
/clone_lib="PPCI-24 Male Mouse BAC"
1..1011
/feature="assembly_fragment"
1112..2019
/feature="assembly_fragment"
2120..2920
/feature="assembly_fragment"
3021..4680
/feature="assembly_fragment"
4781..6502
/feature="assembly_fragment"
6603..8777
/feature="assembly_fragment"
8878..11203
/feature="assembly_fragment"
11304..14001
/feature="assembly_fragment"
14102..16893
/feature="assembly_fragment"
16994..21284
/feature="assembly_fragment"
21385..25083
/feature="assembly_fragment"
25184..33023
/feature="assembly_fragment"
33124..43063
/feature="assembly_fragment"
43164..56342
/feature="assembly_fragment"
56443..75509
/feature="assembly_fragment"
75610..90469
/feature="assembly_fragment"
90570..109549
/feature="assembly_fragment"
109650..160746
/feature="assembly_fragment"
160847..227679
/feature="assembly_fragment"
227780..285533
/feature="assembly_fragment"
285634..324084
/feature="assembly_fragment"
clone_end:T7
vector_side:right"

```

misc_feature

ORIGIN

Query Match 0.9%; Score 39; DB 2; Length 324084;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4333 TGGCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4371

Db 122401 TGGCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 122363

RESULT 56

LOCUS AR418098 138 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 9595 from patent US 6639063.
ACCESSION AR418098

VERSION AR418098.1 GI:40173208
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 138)
AUTHORS Edwards, J.-B.D.M., Jobert, S. and Giordano, J.-Y.
TITLES EST's and encoded human proteins
JOURNAL Patent: US 6639063-A 9595 28-0CT-2003;
FEATURES
source
1. .138
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 0.9%; Score 38; DB 6; Length 138;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAA 4372
Db 68 GCTGTTAAAAA 105

RESULT 57
AX978792 138 bp DNA linear PAT 15-JAN-2004
LOCUS
DEFINITION Sequence 9595 from Patent EPI104808.
ACCESSION AX978792
VERSION AX978792.1 GI:40984932
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Dumas Milne Edwards, J.B., Jobert, S. and Giordano, J.Y.
TITLES ESTs and encoded human proteins
JOURNAL Patent: EP 1104808-A 9595 06-JUN-2001;
FEATURES
source
1. .138
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 0.9%; Score 38; DB 6; Length 138;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAA 4372
Db 68 GCTGTTAAAAA 105

RESULT 58
BD113651 138 bp DNA linear PAT 18-SEP-2002
LOCUS
DEFINITION EST and encoded human protein.
ACCESSION BD113651
VERSION BD113651.1 GI:23208469
KEYWORDS JP 2002010789-A/5728.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Edwards, J.B.D.M., Jobert, S. and Giordano, J.Y.
TITLES EST and encoded human protein
JOURNAL Patent: JP 2002010789-A 5728 15-JAN-2002;
GENSET CORP
COMMENT OS Homo sapiens (human)

PN JP 2002010789-A/5728
PD 15-JAN-2002
PF 07-AUG-2000 JP 2000280989
PR 05-AUG-1999 US 60/147499
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EYE PI
GIORDANO
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC
C12N1/21,
PC C12N5/10, C12P21/02, C12P21/08, C12O1/68, C12N15/00, C12N5/00, PC
C12N15/00
CC EST and encoded human protein
FH Key location/Qualifiers
FT source 1. .138
FT /organism="Homo sapiens (human)"
FT Location/Qualifiers
1. .138
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 0.9%; Score 38; DB 6; Length 138;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAA 4372
Db 68 GCTGTTAAAAA 105

RESULT 59
BC070483 411 bp mRNA linear PRI 11-MAY-2004
LOCUS
DEFINITION Homo sapiens cDNA clone IMAGE:6472970, partial cds.
ACCESSION BC070483
VERSION BC070483.1 GI:47124639
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldi, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramsen, R.D., Mulian, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Munz, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, D.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
REFERENCE
PUBMED
AUTHORS Strausberg, R.
TITLES Direct Submission
JOURNAL Submitted (10-MAY-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk

Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mc@hghri.nih.gov
 Ahlter N., Ayala K., Beckstrom-Sternberg S.M., Benjamin B.,
 Blakesley R.W., Bouffard G.G., Breen K., Brinkley C., Brooks S.,
 Dietrich N.L., Granite S., Guan X., Gupta J., Haghighi P.,
 Hansen N., Ho S.-L., Karlins E., Kwong P., Latic P., Legaspi R.,
 Maduro Q.L., Mastello C., Maskeri B., Mastrian S.D., McCloskey J.C.,
 McDowell J., Pearson R., Stantrop S., Thomas P.J., Touchman J.W.,
 Turgeson C., Vogt J.L., Walker M.A., Wetherby K.D., Wiggins L.,
 Young A., Zhang L.-H. and Green E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>
 Series: IRAC Plate: 174 Row: 0 Column: 20
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 10863874.
 Location/Qualifiers
 1. 411
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6472970"
 /tissue_type="uterus", leiomyosarcoma"
 /clone_id="N1H MGC_71"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN
 Query Match 0.9%; Score 38; DB 9; Length 411;
 Best Local Similarity 100.0%; Pred. No. 3.8e-10;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 4335 GCTGTTAAAAA 4372
 Db 368 GCTGTTAAAAA 405

RESULT 60
 BC074531 1347 bp mRNA linear VRT 20-SEP-2004
 LOCUS
 DEFINITION
 XENOPUS TROPICALIS MGC69355 protein, mRNA (cDNA clone MGC:69355
 IMAGE:5309295), complete cds.
 BC074531
 BC074531.1 GI:49257721
 MGC.
 Xenopus tropicalis (Silurana tropicalis)
 Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 1347)
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 Datchenko L., Marusina K., Farmer A.A., Rudin G.M., Hong L.,
 Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L.,
 Scheetz T.E., Brownstein M.J., Uddin T.B., Toshiyuki S.,
 Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J.,
 Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J.,
 McKernan K.J., Malek J.A., Gunaratne P.H., Richards S.,
 Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Vallalath D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S.,
 Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y.,
 Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D.,

Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 Butterfield Y.S., Krzywinski M.I., Skalska U., Smalios D.B.,
 Scherch A., Schein J.E., Jones S.J. and Marra M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 1347)
 Klein S. and Gerhard D.S.
 JOURNAL
 DIRECT SUBMISSION
 Submitted (23-JUN-2004) National Institutes of Health, Xenopus Gene
 Collection (XGC), National Institute of Child Health and Human
 Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
 20892-7510, USA
 NIH-MGC Project
 Contact: XGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Robert M. Grainger
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: DOE Joint Genome Institute, Walnut Creek,
 California
 Web site: www.jgi.doe.gov

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>
 Series: IRAC Plate: 130 Row: 0 Column: 22
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, similarity but not identity to protein.
 Location/Qualifiers
 1. 1347
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="MGC:69355 IMAGE:5309295"
 /tissue_type="Embryo", Silurana tropicalis, gastrula
 (stages 10-13)"
 /clone_id="N1CHD_XGC_Emb5"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6.1"
 1. 1347
 /gene="MGC69355"
 /db_xref="LOCUSID:448015"
 129. 806
 /gene="MGC69355"
 /codon_start=1
 /product="MGC69355 protein"
 /protein_id="AAH74531.1"
 /db_xref="GI:49257722"
 /db_xref="LOCUSID:448015"
 /translation="MSGMDREAAKTSGGGRIVGNLPADIREKELEDLPDRIRIT
 IEKRNKGSSAAPFAFISFODPRDAEAVARNGEYGSRLRVEPFRSGSGGSG
 GSRGRNPGPSPRSREYIVIGLPPSGMOQLKHMRAAGVCYADVHKDGMGIVELFR
 KEMEEVALRLDITKFRSHGETSYIVCPERNITSYRSRSRSGRDSVPQSRSPRY
 ASPPRPY"

ORIGIN
 Query Match 0.9%; Score 38; DB 5; Length 1347;
 Best Local Similarity 100.0%; Pred. No. 4.7e-10;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 4335 GCTGTTAAAAA 4372
 Db 1290 GCTGTTAAAAA 1327

RESULT 61
 ZMUG6404 1459 bp mRNA linear PLN 14-JAN-1997
 LOCUS
 DEFINITION
 Zea mays plastid phosphate/phosphoenolpyruvate translocator
 precursor (MZPP74) mRNA, complete cds.
 ZMUG6404
 ZMUG6404.1 U66404

VERSION U66404.1 GI:1778148
KEYWORDS
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.
AUTHORS 1 (bases 1 to 1459)
TITLE Fischer, K., Weber, A. and Fluegge, U. I.
JOURNAL The phosphoenolpyruvate translocator from maize plastids
REFERENCE 2 (bases 1 to 1459)
AUTHORS Fischer, K., Weber, A. and Fluegge, U. I.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-1996) Botany II, University of Cologne, Gyrhofstrasse 15, Koeln 50931, Germany
FEATURES
source
1. 1459
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
1. 1459
/gene="MZPPT4"
26. 1216
/gene="MZPPT4"
/function="catalyzes the transport of phosphoenolpyruvate and phosphate across the inner envelope membrane of plastids"
/note="PPT"
/codon_start=1
/product="phosphate/phosphoenolpyruvate translocator precursor"
/protein_id="AA040650.1"
/db_xref="GI:1778149"
/translation="MOSAAALGVPCARPLPTYPSPRRGAGCSGAQPIIAPRGL CTSARASILPASPLEEYRCRAAGTCRKGVAADGAVEAGGLVKTQLSLFGL WYLFYFNTYVYNKQVLPYINITTVQVGSATIALFWITGILKRPTISAOJLA IPIAIVHTMGNTFTNMSLQKAVSPHTTKAMPSPSVLSAIFGELPTPVVLSL LPIVGGVALASLTESAFNMGFVMSANSTFOSRNVLSKKLWKKESLDNNLSI ITWSPFLAPVTLTEGVASPAVLQSAELNKQIYTRSLAACCFHAYQVSYML ARVSPYTHSVGNCVKRVVIVTSVLFPRTPVSINSIGTIALAGFLYSQLRLMK PKAA"

ORIGIN
Query Match 0.9%; Score 38; DB 8; Length 1459;
Best Local Similarity 100.0%; Pred. No. 4.8e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4335 GCTGTTAAAAA... 4372
1412 GCTGTTAAAAA... 1449

RESULT 62
LOCUS BC024496 1482 bp mRNA linear ROD 06-OCT-2003
DEFINITION Mus musculus RIKEN cDNA 1810019C21 gene, mRNA (cDNA clone MGC:37431 IMAGE:4982244), complete cds.
ACCESSION BC024496
VERSION BC024496.1 GI:19353580
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 1482)
Krausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, J., Stenmen, C. M., Schuler, G. D., Altschul, S. F., Zeeberg, B., Bueltow, K. H., Schaefer, C. F., Bhat, N. K., Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Heide, F., Datchenko, L., Marusina, K., Farmer, A. R., Rubin, G. M., Hong, L., Stapleton, M., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Brownstein, M. J., Udén, T. B., Toshiyuki, S.,

REMARK
COMMENT
JOURNAL MEDLINE
PUBMED 22388257
12477932
TITLE 2 (bases 1 to 1482)
AUTHORS Strausberg, R.
JOURNAL Direct Submission
Submitted (01-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) med@paki.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAK Plate: 59 Row: m Column: 22
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21703901.
FEATURES
source
1. 1482
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:37431 IMAGE:4982244"
/tissue_type="Colon, normal, 5 month old male mouse."
/clone_id="NCT CGAP_C024"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. 1482
/gene="1810019C21R1k"
/db_xref="locustid:69085"
/db_xref="MGI:1916335"
38. 859
/gene="1810019C21R1k"
/codon_start=1
/product="1810019C21R1k protein"
/protein_id="AA024496.1"
/db_xref="GI:19353581"
/db_xref="MGI:1916335"
/translat="MTRMARVVTTSNSKRPISATSWEDMKGSVERADQSLPNKOCOS SRPLRNDSPQARKKKKKKYEYINADVNGPMFYLTQNSQVLEHGOIADDSQVEBC IYALAKDSRBERRLKROAKKNAVCHGPGGIADCPAVYESQDMGTGICVRC GSTHEHMSKRAVNDPALGFPPAKFCVGEKHLRS CPDNTKGYADGCSCKLCS VEPFKDQRENQSDRIITVGRWAKGMSADYEDVDVPLKQKPTKVPKVNPLY"
320. 742
/gene="1810019C21R1k"
/note="AIR1: Region: C0G5082, AIR1, Arginine methyltransferase-interacting protein, contains RING

Zn-finger [Posttranslational modification, protein turnover, chaperones / Intracellular trafficking and secretion]
 /db_xref="CDD:COG5082"

ORIGIN
 Query Match 0.9%; Score 38; DB 10; Length 1482;
 Best Local Similarity 100.0%; Pred. No. 4.9e-10;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
 1445 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1482

RESULT 63
 BC004226 1590 bp mRNA linear PRI 16-SEP-2003
 LOCUS Homo sapiens regulator of G-protein signalling 19 interacting
 DEFINITION protein 1, mRNA (cDNA clone IMAGE:3538778), partial cds.
 ACCESSION BC004226
 VERSION BC004226.2 GI:33872739
 KEYWORDS
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1590)
 Klausner,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Bueltow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heileh,F.,
 Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Shapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Morley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Vallalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shvachenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butterfield,Y.S., Krzywinski,M.T., Skalska,U., Smalins,D.E.,
 Schmeich,A., Schein,J.B., Jones,S.J., and Marz,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL
 MEDLINE
 PUBMED 22388257
 REFERENCE 12477932
 AUTHORS 2 (bases 1 to 1590)
 TITLE Strausberg,R.
 JOURNAL Direct Submission
 Submitted (01-MAR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 On Aug 19, 2003 this sequence version replaced gi:13278947.
 Contact: MGC help desk
 Email: gcgaps-remail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 contact: amadamsystemsbio.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha
 Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov

FEATURES
 source
 Series: IRAL Plate: 11 Row: m Column: 16.
 Location/Qualifiers
 1..1590
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3538778"
 /tissue_type="lung, small cell carcinoma"
 /clone_lib="NIH MGC 7"
 /lab_host="DH10B-R"
 /note="vector: pOTB7"
 <1..1590
 /gene="RGS191P1"
 /note="synonyms: GIPC, TIP-2, GLUT1CBP, IIP-1, C19orf3,
 SEMCAP, NIP, SYNECTIN, Hs.6454"
 /db_xref="tlocusid:10755"
 /db_xref="MIM:605072"
 <1..904
 /gene="RGS191P1"
 /codon_start=2
 /product="RGS191P1 protein"
 /protein_id="AAH04226.2"
 /db_xref="GI:33872740"
 /db_xref="MIM:10755"
 /db_xref="MIM:605072"
 /translation="PIGGGGGGPQWGLPEPPALRPVFTHTOLAHGSPTRLEGFT
 NVKEIKGIAEAFRLPTAEYMFCTLNTHTKMDKILGGQIGLEDFIAHYKGRKEVE
 VFSEDAIGITTDNGAGYAFIRIGESYTDIHLISVDMTEALNGSLGCRHE
 VARLKEIPRGFTTLKTEPRAPFMSISRGSRGSPQGTGTURLSRGPA
 TVDLPASAEKKAIEKVDLESYMGIRDTETLAATWELGDKRNPDELAALDERLG
 DPAFPDEFPVDMGAIADKAGRY"
 290..532
 /gene="RGS191P1"
 /note="PDZ; Region: Domain present in PSD-95, Dlg, and
 ZO-1/2. Also called DHR (Dlg homologous region) or GUGF
 (relatively well conserved tetrapeptide in these domains).
 Some PDZs have been shown to bind C-terminal polypeptides"
 /db_xref="CDD:smart00228"

ORIGIN
 Query Match 0.9%; Score 38; DB 9; Length 1590;
 Best Local Similarity 100.0%; Pred. No. 4.9e-10;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
 Db 1553 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1590

RESULT 64
 BC002914 1680 bp mRNA linear PRI 04-OCT-2003
 LOCUS Homo sapiens Wiskott-Aldrich syndrome protein interacting protein,
 DEFINITION mRNA (cDNA clone MGC:10429 IMAGE:3954968), complete cds.
 ACCESSION BC002914
 VERSION BC002914.2 GI:33877341
 KEYWORDS MGC.
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1680)
 Klausner,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Bueltow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heileh,F.,
 Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Shapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Morley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,

RESULT 67
BC012805
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BC012805 2586 bp mRNA linear PRI 25-MAR-2004
Homo sapiens RAN binding protein 6, mRNA (cDNA clone
IMAGE:3507184), complete cds.
BC012805
BC012805.2 GI:38014037
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2586)
Strausberg,R.L., Feigold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heish,F.,
Datchenko,L., Marsina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmittz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalski,U., Smalins,D.E.,
Scherer,A., Schein,J.E., Jones,S.J. and Marra,M.A.
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2586)
Strausberg,R.
Direct Submision
Submitted (15-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Oct 28, 2003 this sequence version replaced gi:15277474.
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbiology.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAL Plate: 7 Row: a Column: 14
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: The cds is short compared to
the longest cds in the locus.
Location/Qualifiers
1..2586
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3507184"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH_MGC_14"
/lab_host="DH10B-R"

gene
CDS
/note="Vector: pOTB7"
1..2586
/gene="RANBP6"
/db_xref="LocustID:26953"
47..1291
/gene="RANBP6"
/codon_start=1
/product="RANBP6 protein"
/protein_id="AAH12805.2"
/db_xref="GI:31416798"
/db_xref="LocustID:26953"
/translation="MLVYAKELRGFVETROYKLVPLKPYPHNDVRAAERSM
PFLBGRIRGPEYLAQMOPICDPLIKATIGREPTDYSETMNSAKSLYVMGOST
NDHSEBEGGILKALBGFKNQELRQRESENPDQVQKSLQDECDVITLTVS
DILHSIFSTYKELLPWEQLPLIVNLICSSRPMPDRQGLCTFDIILIEHCSPTSEK
VYEYFPMPLLMNRNDPEVRQAAAYGLGVAGFGGDYRSICSEAVPLLVICAN
STKRWVILATENCISAIKILFKPCVAVDEVLPMTLSMLPLHEDKEKAIOTLSFC
DIESNHPVPIGIPNNNSLPIKIIIAEGKINNTINVEDPCARLANVAVQVTSBDM
LRCVSLDDEQDAQLINFA"

ORIGIN
Query Match 0.9%; Score 38; DB 9; Length 2586;
Best Local Similarity 100.0%; Pred. No. 5.3e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4335 GCTGTTAAAAAAGAAAAAAGAAAAAAGAAAAA 4372
DB 2532 GCTGTTAAAAAAGAAAAAAGAAAAAAGAAAAA 2569

RESULT 68
BC035618
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BC035618 4000 bp mRNA linear PRI 29-JUN-2004
Homo sapiens somatostatin receptor 1, mRNA (cDNA clone MGC:45280
IMAGE:5175220), complete cds.
BC035618
BC035618.1 GI:23272573
MGC.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 4000)
Strausberg,R.L., Feigold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heish,F.,
Datchenko,L., Marsina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmittz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalski,U., Smalins,D.E.,
Scherer,A., Schein,J.E., Jones,S.J. and Marra,M.A.
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 4000)
Strausberg,R.
Direct Submision
Submitted (31-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)
Gaithersburg, Maryland
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@hghri.nih.gov

Ahter, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, O.L., Masello, C., Maekeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantrop, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: <http://image.llnl.gov>
Series: IRAC Plate: 68 Row: 1 Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4557856.
Location/Qualifiers
1. 4000
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:45280 IMAGE:5175220"
/tissue_type="Brain, Lung, Testis, adult, pooled whole"
/clone_id="NIM MGC_115"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. 4000
/gene="SSTR1"
/note="synonym: SRIF-2"
/db_xref="LocusID:6751"
/db_xref="MIM:182451"
275. 1450
/gene="SSTR1"
/codon_start=1
/product="somatostatin receptor 1"
/protein_id="AAH35618.1"
/db_xref="GI:23272574"
/db_xref="LocusID:6751"
/db_xref="MIM:182451"
/translation="MPNNGTASSPSSPSPGSCRGGSGRGAGAGADGMEBPGRN
ASONGTLESGQSAIILISFYVCLVGLGNSVYIYLIRAKMTATNIIYLNLAI
ADBLMLSVPLVTSSTLRHMPGALICRLVLSVDNMTSTYCLTVLSVDRYVAIV
HPIKAAVRRPTVAKVYVNLGVMLSLVTLPIVFFSTANSGTVA CNMLMBPAPR
WLVGFVLTFLMGLFLPVGAICICVYIIIMKRNVALKAGMOGRKSERKITLVMVW
VMVFVICMPPYVQVAVNFAEDDADATVSQLSVLIGVANS CANPILYGFISDNPKSF
ORILCLSWMDNAAEPPVYATALKSRVSVEDFQPNLESGVFRNGTCTSRITTL"

ORIGIN
Query Match 0.9%; Score 38; DB 9; Length 4000;
Best Local Similarity 100.0%; Pred. No. 5.7e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4335 GCTGTTAAAAA 4372
Db 3943 GCTGTTAAAAA 3980

RESULT 69
AL928696 102448 bp DNA linear ROD 21-NOV-2002
LOCUS Zebrafish DNA sequence from clone Rp23-387G11 on chromosome 2, complete
DEFINITION Mouse DNA sequence from clone Rp23-387G11 on chromosome 2, complete
ACCESSION AL928696
VERSION AL928696.6 GI:25168716
KEYWORDS HTG.

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 102448)
TITLE Sycamore, N.
JOURNAL Direct Submission
Submitted (21-NOV-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Nov 22, 2002 this sequence version replaced gi:25136705.
Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30) or an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep>
from the RPI-23 Mouse PAC Library
constructed by the group of Peter de Jong.
For further details see <http://www.choxi.org/bacpac/home.htm>
VECTOR: pBACe3.6.
Location/Qualifiers
1. 102448
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-387G11"
/clone_id="RPI-23"

ORIGIN
Query Match 0.9%; Score 38; DB 10; Length 102448;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4335 GCTGTTAAAAA 4372
Db 93169 GCTGTTAAAAA 93206

RESULT 70
BX664613 121435 bp DNA linear VRT 19-APR-2004
LOCUS Zebrafish DNA sequence from clone DKEY-5P5 in linkage group 22, complete sequence.
DEFINITION BX664613
ACCESSION BX664613.17 GI:46406619
VERSION BX664613.17
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio

REFERENCE
AUTHORS
TITLE
JOURNAL

Wright, D.
Submitted (17-APR-2004), Wellcome Trust Sanger Institute, Hinxton.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep

Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Racoon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_reio/fishmask.shtml

DKEY-595 is from a Zebrafish AC library

VECTOR: pindigBAC-5.

ORIGIN

Query Match	0.9%;	Score 38;	DB 5;	Length 121435;
Best Local Similarity	100.0%;	Pred. No. 1e-09;		

OY	4335	GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAA	4372
Db	54499	GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAA	54622

Db 54499 GCTGTTAAAAA 54462

RESULT 71	
AL354659	
LOCUS	135060 bp DNA linear PRI 15-NOV-2001
DEFINITION	Human DNA sequence from clone RPL1-157G15 on chromosome 1, complete sequence.
ACCESSION	AL354659 AC024007
VERSION	AL354659.14 GI:16972840
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)

ORGANISM *Homo sapiens*
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 135060)
Griffiths, C.
Direct Submission
Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton.

IMPORTANT: This sequence is not the entire insert of clone RP11-157G15. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone RP11-157G15 is at 135060 in this sequence. The true left end of clone RP11-291M23 is at 34561 in this sequence. The true right end of clone RP11-361K17 is at 2000 in this sequence.

ORIGIN

Query Match	0.9%;	Score 38;	DB 9;	Length 135060;
Best Local Similarity	100.0%;	Pred. No. 1.1e-09;		
Matches 30	0	0	0	0
Genes	0	0	0	0
MicroRNAs	0	0	0	0

[illegible]

Db 126281 CTTTGGTTTATGCCCATTTTCATATTGTTGTCGTGTTG 126318

RESULT	72
AL513172	
LOCUS	142908 bp DNA linear HTG 10-JUL-2001
DEFINITION	Homo sapiens chromosome 1 clone RP11-229H23, 17 unordered pieces.
ACCESSION	AL513172
VERSION	AL513172.1 GI:12539823
KEYWORDS	HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 McIay,K.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
----- Genome Centre
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA291H23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 137540 bases at least Q40
Consensus quality: 139579 bases at least Q30
Consensus quality: 140684 bases at least Q20
Insert size: 141308; sum-of-contigs
Insert size: 162784; 14.9% error; agarose-fp
Quality coverage: 4.48x in Q20 bases; sum-of-contigs Quality
Coverage: 4.05x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 11819: contig of 11819 bp in length
11820 11919: gap of 100 bp
11920 21066: contig of 9147 bp in length
21067 21166: gap of 100 bp
21167 32577: contig of 11411 bp in length
32578 32677: gap of 100 bp
32678 35204: contig of 2527 bp in length
35205 35304: gap of 100 bp
35305 38616: contig of 3312 bp in length
38617 38716: gap of 100 bp
38717 48887: contig of 10171 bp in length
48888 48987: gap of 100 bp
48988 51811: contig of 2824 bp in length
51812 51911: gap of 100 bp
51912 54004: contig of 2093 bp in length
54005 54104: gap of 100 bp
54105 60083: contig of 5979 bp in length
60084 60183: gap of 100 bp
60184 68204: contig of 8021 bp in length
68205 68304: gap of 100 bp
68305 72836: contig of 4532 bp in length
72837 72936: gap of 100 bp
72937 88769: contig of 15833 bp in length
88770 88869: gap of 100 bp
88870 98455: contig of 9586 bp in length
98456 98555: gap of 100 bp
98556 102201: contig of 3646 bp in length
102202 102301: gap of 100 bp
102302 128671: contig of 26370 bp in length
128672 128771: gap of 100 bp
128772 131983: contig of 3212 bp in length
131984 132083: gap of 100 bp
132084 142908: contig of 10825 bp in length.
Location/Qualifiers
1. 142908
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-291H23"

misc_feature /clone_11b="RP11-11.2"
1. 11819
/note="assembly_fragment:00446
clone_end:SP6
vector_side:left"
11920..21066
/note="assembly_fragment:01125
fragment_chain:1"
21167..32577
/note="assembly_fragment:01044
fragment_chain:1"
32678..35204
/note="assembly_fragment:01155
fragment_chain:1"
35305..38616
/note="assembly_fragment:00076
fragment_chain:1"
38717..48887
/note="assembly_fragment:01431
fragment_chain:2"
48988..51811
/note="assembly_fragment:01037
fragment_chain:2"
51912..54004
/note="assembly_fragment:01067
fragment_chain:2"
54105..60083
/note="assembly_fragment:01281
fragment_chain:2"
60184..68204
/note="assembly_fragment:00264
fragment_chain:3"
68305..72836
/note="assembly_fragment:00682
fragment_chain:3"
72937..88769
/note="assembly_fragment:00228"
88870..98455
/note="assembly_fragment:00556"
98556..102201
/note="assembly_fragment:00790"
102302..128671
/note="assembly_fragment:01503"
128772..131983
/note="assembly_fragment:01404
fragment_chain:4"
132084..142908
/note="assembly_fragment:00926
fragment_chain:4
clone_end:T7
vector_side:right"
ORIGIN
Query Match 0.9%; Score 38; DB 2; Length 142908;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4205 CTTTGTATTATGCCCATTTCAATTTGTTGTCGTGTTG 4242
|||||
Db 1231778 CTTTGTATTATGCCCATTTCAATTTGTTGTCGTGTTG 123815
|||||
RESULT 73
AC133929/c AC133929 152549 bp DNA linear HTG 19-OCT-2002
LOCUS Sus scrofa clone RP44-85M9, WORKING DRAFT SEQUENCE, 4 ordered
DEFINITION
pieces.
ACCESSION AC133929 GI:24137428
VERSION AC133929.2
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Center clone name: RP11-415124 (sc0692)

----- Summary Statistics -----

Sequencing vector: plasmid; 62% of reads
Chemistry: Dye-terminator ET; 92% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165135 bases at least Q40
Consensus quality: 165301 bases at least Q30
Consensus quality: 165311 bases at least Q20
Insert size: 165311; sum-of-coverage
Quality coverage: 10.5x in Q20 bases; sum-of-coverage

----- Overlapping Sequences -----

5': Mapping in progress
3': Mapping in progress

----- Sequence Quality Assessment -----

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation -----

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

ECORI HindIII BglIII

SeqDerMap	FngRPrint	SeqDerMap	FngRPrint	SeqDerMap	FngRPrint
8696	8888	2946	2958	3938	3943
6	<800	6382	6573	2067	2063
1114	1095	512	<800	6223	5919
2678	2839	449	<800	3058	3059
12640	12720	1478	1474	1596	1675
3027	3001	4425	4355	1864	1854
3092	3205	4873	4812	2371	2419
3033	3001	4854	4812	2950	3177
2024	2026	101	<800	66	<800
5868	5803	6408	6573	1920	1854
2058	2026	917	912	8892	8587

5042	4888	650	<800	3816	3591
4032	4019	13	<800	1133	1206
3619	3528	1690	1726	4202	4243
591	<800	386	<800	3573	3372
4588	4489	2804	2844	1849	1854
2408	2424	4197	4127	5100	4972
7317	7290	1498	1474	4855	4668
894	900	2333	2322	786	738
3436	3528	18208	18271	2637	2661
628	<800	2298	2322	1590	1498
4548	4489	655	<800	1551	1498
10022	9975	4722	4639	2404	2419
1294	1270	3350	3349	2367	2419
3368	3404	1106	1120	7404	7763
2605	2636	6	<800	2550	2661
11049	11044	1252	1252	5821	5919
2901	3001	1865	1874	6442	6510
3810	3763	3820	3748	7628	7763
3224	3404	3738	3748	3409	3372
265	<800	1260	1252	5101	5174
1506	1489	2177	2158	4674	4668
2794	3001	11327	11212	1871	1854
10596	10633	8132	8134	3928	3943
5955	5803	8234	8134	3151	3177
2975	3001	9110	9098	10946	10956
7368	7290	7551	7596	1507	1498
740	<800	1891	1874	5891	5919
6924	6877	3698	3748	8949	9128
488	<800	218	<800	2678	2827
1689	1615	11309	11212	51	<800
634	<800	770	<800	6768	6510
376	<800	8011	8134	3884	3943
12088	12061	581	<800	2500	2661
		1780	1874	4918	4800
		641	<800	719	<800
		5216	5119	2412	2419

```

2255      2322
-----
1179      1120
-----
734      <800
-----
Location/Qualifiers
1..165311
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-415L24"
/clone_1ib="RP11 human BAC library 11"

ORIGIN
Query Match      0.9%; Score 38; DB 9; Length 165311;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4205 CTTTGTATTATGCCCATTTTCATTTGTGTCTGTGTTG 4242
          |||||||
Db      21531 CTTTGTATTATGCCCATTTTCATTTGTGTCTGTGTTG 21568

RESULT 75
AL928975      174480 bp      DNA      linear      VRT 14-MAR-2003
LOCUS      Zebrafish DNA sequence from clone CH211-133N4, complete sequence.
DEFINITION      AL928975
ACCESSION      AL928975
VERSION      AL928975.10 GI:28971544
KEYWORDS      HTG.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
              1 (bases 1 to 174480)
              Almeida,J.
              Direct Submission
              Submitted (14-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              zfish-help@sanger.ac.uk Clone requests:clonerequest@sanger.ac.uk
              On Mar 14, 2003 this sequence version replaced GI:28445862.
              ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: zfish-help@sanger.ac.uk
              -----

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhiron Bao and Sean Eddy, submitted), and those beginning 'dr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www/Projects/D_rerio/fishmask.shtml

```

CH211-133N4 is from a CHORI-211 BAC library
VECTOR: pTRBAC2.1.
Location/Qualifiers
source
1..174480
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-133N4"
/clone_1ib="CHORI-211"

ORIGIN
Query Match      0.9%; Score 38; DB 5; Length 174480;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4333 TGGCTGTATTAATAAAAAAAAAAAAAAAAAAAAAA 4370
          |||||||
Db      102442 TGGCTGTATTAATAAAAAAAAAAAAAAAAAAAAAA 102405

Search completed: March 25, 2005, 10:44:35
Job time : 18325 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2005, 05:02:49 ; Search time 13100 Seconds
(without alignments)
12703.587 Million cell updates/sec

Title: US-10-054-935-1

Perfect score: 4372

Sequence: 1 cagctcctgaccccccacac.....aaaaaaaaaaaaaaaaaaaa 4372

Scoring table: OLIGO_NUC

Searched: 34239544 seqs, 19032134700 residues

Word size : 15

Total number of hits satisfying chosen parameters: 3767501

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3443	78.8	3785	3	CR749360 Homo sapi
2	776	17.7	927	5	B0857520 AGENCOURT
3	741	16.9	1056	4	BM547318 AGENCOURT
4	734	16.8	769	7	CN355147 AGENCOURT
5	711	16.3	729	5	BQ004676 UI-H-E10-
6	705	16.1	743	7	CN355143 170006001
7	703	16.1	765	5	B0014717 UI-H-E10-
8	698	16.0	931	5	BQ948768 AGENCOURT
9	697	15.9	637	4	B0682595 UI-CF-BE1
10	693	15.9	817	4	BI917709 603183689
11	689	15.8	704	7	CN355144 170005326
12	678	15.5	870	5	BQ229745 AGENCOURT
13	673	15.4	888	5	B0517114 AGENCOURT
14	668	15.3	792	5	B0929398 AGENCOURT
15	664	15.2	715	5	BM981391 UI-CF-BE1
16	653	14.9	712	6	CA447821 UI-H-E10-
17	642	14.7	921	5	B0160445 AGENCOURT
18	639	14.6	1032	4	BM475963 AGENCOURT
19	637	14.6	891	4	B0898842 AGENCOURT
20	632	14.5	789	6	CD652041 AGENCOURT
21	623	14.2	640	6	CA307817 AGENCOURT
22	622	14.2	896	6	BM329236 BX329236
23	621	14.2	902	5	B0517052 AGENCOURT
24	615	14.1	615	7	CK430007 of43h12.y

25	613	14.0	640	2	BE896063
26	613	14.0	641	7	CN355145
27	611	14.0	891	5	BQ881177 AGENCOURT
28	607	13.9	756	6	CB851322 UI-CF-BE1
29	601	13.7	601	6	CD674654 f812901.y
30	597	13.7	745	2	BE907276
31	594	13.6	611	2	BE693636
32	593	13.6	644	5	BX952313
33	587	13.4	716	6	CD639095
34	586	13.4	593	5	B0781747
35	576	13.2	697	4	BI820795
36	570	13.0	864	4	BQ289135
37	561	12.8	723	5	BQ187281
38	558	12.7	558	2	AM182868
39	556	12.7	621	4	BI032095
40	553	12.6	629	1	AV722833
41	551	12.6	617	4	BM770559
42	551	12.6	628	7	CK004541
43	546	12.5	643	6	CA447820 UI-H-E10-
44	544	12.4	544	5	BM989342 UI-H-DP0-
45	542	12.4	567	4	BM729196
46	540	12.4	540	4	BM852930
47	536	12.3	536	1	AA908959
48	536	12.3	536	4	BG547204
49	536	12.3	536	4	BM754901
50	536	12.3	623	4	BI457952
51	533	12.2	608	5	B0636732
52	533	12.2	721	7	CK300608
53	532	12.2	666	2	AM965141
54	531	12.1	531	2	BE870003
55	530	12.1	753	6	CA313331
56	528	12.1	696	6	CA444700
57	527	12.1	527	4	BM661885
58	525	12.0	525	1	AI628880
59	524	12.0	524	2	AM173363
60	524	12.0	529	4	BM711612
61	524	12.0	533	4	BM665523
62	523	12.0	541	4	BM264268
63	522	11.9	522	2	AM665845
64	518	11.8	770	5	BX479913
65	515	11.8	710	1	AV713546
66	506	11.6	506	7	CR738569
67	502	11.5	985	5	BM970469
68	499	11.4	686	5	BM970469
69	498	11.4	549	4	AI800794
70	497	11.4	951	4	BG178033
71	495	11.3	551	4	BI667371
72	492	11.3	508	5	BQ127983
73	492	11.3	510	7	W03542
74	491	11.2	491	1	AA746655
75	489	11.2	523	1	AI151418
76	474	10.8	474	1	AI093113
77	473	10.8	473	1	AI137796
78	473	10.8	758	2	BE1792962
79	472	10.8	472	1	AI624029
80	472	10.8	478	1	AI635625
81	472	10.8	479	1	AI951984
82	469	10.7	700	1	AU252433
83	468	10.7	648	6	CB321703
84	466	10.7	466	1	AI1346936
85	466	10.7	524	4	BM689858
86	464	10.6	585	4	BI019744
87	464	10.6	1226	5	BM904456
88	462	10.6	463	1	AI954139
89	460	10.5	460	1	AI582361
90	460	10.5	531	5	BM969790
91	459	10.5	459	7	W92652
92	459	10.5	518	7	CF138377
93	455	10.4	506	1	AI653663
94	455	10.4	753	6	CD356812
95	451	10.3	451	7	AI831057
96	449	10.3	570	7	CN355142
97	449	10.3	700	1	AU252432

BE896063	601438862
CN355145	170006000
BQ881177	AGENCOURT
CB851322	UI-CF-BE1
CD674654	f812901.y
BE907276	601500975
BE693636	MR4-BT035
BX952313	DKFZP781M
CD639095	AGENCOURT
B0781747	im94e12.y
BI820795	603034002
BQ289135	602387523
BQ187281	UI-E-E10-
AM182868	xj54h09.x
BI032095	PM4-MT042
AV722833	K-EST0054
BM770559	K-EST0054
CK004541	AGENCOURT
CA447820	UI-H-E10-
BM989342	UI-H-DP0-
BM729196	UI-E-E01-
BM852930	K-EST0134
AA908959	cl10h04.6
BC547204	602574616
BM754901	K-EST0032
BI457952	603198945
B0636732	h413906.y
CK300608	UI-E-E10-
AM965141	EST377214
BE870003	601449751
CA313331	UI-CF-FNO
CA444700	UI-H-DH1-
BM661885	UI-E-CI1-
AI628880	ly95f12.x
AM173363	xj78e12.x
BM711612	UI-E-CI1-
BM665523	UI-E-CL1-
BM264268	1933b05.y
AM665845	h194f04.x
BX479913	DKFZP686F
AV713546	AV713546
CR738569	CR738569
BM970469	UI-CF-BE1
AI800794	w913c04.x
BG178033	602327595
BI667371	603292145
BQ127983	1j79g08.y
W03542	y29p09.y
AA746655	nx27d10.s
AI151418	qc88e08.x
AI093113	qa87c07.x
AI137796	t6i1b02.x
BE1792962	601585622
AI624029	t825g10.x
AI635625	t220h11.x
AI951984	wx40c05.x
AU252433	AU252433
CB321703	UI-CF-BE1
AI1346936	qp54a08.x
BM689858	UI-E-CK1-
BI019744	1l3-MT026
BM904456	AGENCOURT
AI954139	wx80b09.x
AI582361	t6g7b10.x
BM969790	UI-CF-DU1
W92652	ze04g03.x
CF138377	UI-HF-BNO
AI653663	t221b02.x
CD356812	AGENCOURT
AI831057	wj62d01.x
CN355142	170005325
AU252432	AU252432

```
98      449  10.3  705  2  BE277857  601120091
      497  10.2  560  2  BX478274  DKFZp686J1222
      100  447  10.2  841  4  B1517294  B1517294  603041522

ALIGNMENTS

RESULT 1
CR749360      3785 bp      mRNA      linear      HTC 19-AUG-2004
LOCUS      Homo sapiens mRNA; cDNA DKFZp686J1211 (from clone DKFZp686J1211).
DEFINITION      CR749360
ACCESSION      CR749360.1  GI:51476445
KEYWORDS      HTC.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 3785)
Koehler,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Ossanger,A.,
Fobo,G., Han,M. and Wiemann,S.
The German CDNA Consortium
Direct Submission
Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuberberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the
Heinrich-Heine-University, Dueseldorf/Germany) within the CDNA
sequencing consortium of the German Genome Project. This clone
(DKFZp686J1211) is available at the RZPD Deutches
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp686J1211
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.

FEATURES
source      location/Qualifiers
1..3785
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686J1211"
/issue_type="rectum tumor"
/clone_lib="686 (synonym: hlcc03). Vector pSport1_Sfi; host
Dh10B; Bates SfiIA + SfiIB"
/dev_stage="adult"
/note="hypothetical protein, differentially spliced"
1..3785
/gene="DKFZp686J1211"
107..1348
/gene="DKFZp686J1211"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH18213.1"
/db_xref="GI:51476446"
/translaction="MRKSPGSGSGSGSQAACLQILLQLDLIEQQQQLQAKX
EIEBELSERDITLLARIERMRQVLVKONKERNRLFOYETREBEETLSKIKLE
COPELSETQTLPPKPGSGSGSKHRSKSPGSTRKTPVKTALEFSEIVSKIKLE
SPIKERPGSLSETVCKRELRSQETPEKRSASVDPRLSTPKGSGTHPKKAPSS
IEDLPILPTTEMTLCRNHQPSPPLRBSRPKKESTVARCLMPSSVAGSTVLAPPS
WEDHSVPLRDNPSDLLENLSDVSFKRAKULEKRRKRDILRIRORLLOLQ
LMWKKKGIQESPEVTSFPPEPDVESLMTTPFLPVAFGRLPLRLDONFELPILD
ERSRRLRIQKOTPHRTCRK"
```

ORIGIN

```
Query Match      78.8%; Score 3443; DB 3; Length 3785;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3693; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY      662  CGAGGCGAGATGAGATGAGTATGAGAAAGAGCCCTCTCGGGGATGATGCGGCTCGGAGC 721
      |||||
      DB      88  CGAGGCGAGATGAGATGAGTATGAGAAAGAGCCCTCTCGGGGATGATGCGGCTCGGAGC 147
```

```
QY      722  CTCGAGTCAGGCGCGCTGCTGCTCAACAGATCCTTCTGCTGCAATTGGACCTCATCGAAC 781
      |||||
      DB      148  CTCGAGTCAGGCGCGCTGCTGCTCAACAGATCCTTCTGCTGCAATTGGACCTCATCGAAC 207

QY      782  GCAGCAGCAGCAGCTGCGAGCCAGAGAAAGAGATCGAGAGCTGAAGTCAGAGAGAG 841
      |||||
      DB      208  GCAGCAGCAGCAGCTGCGAGCCAGAGAAAGAGATCGAGAGCTGAAGTCAGAGAGAG 267

QY      842  CACGCTCTTGCTCGGATTGAACGTATGGAAGGCGGATGACAGCTGTATAAGAGATTA 901
      |||||
      DB      268  CACGCTCTTGCTCGGATTGAACGTATGGAAGGCGGATGACAGCTGTATAAGAGATTA 327

QY      902  CGAGAAAGAAAGCAGCAAGCTGTTTCAAGGCTATGAAACTGGAAGAGAGAGAAACAG 961
      |||||
      DB      328  CGAGAAAGAAAGCAGCAAGCTGTTTCAAGGCTATGAAACTGGAAGAGAGAGAAACAG 387

QY      962  GCTATCTGAGAAATTAACCTGAGAGCCAGCGAGCTTTCCGAGACATCCGAGACTCT 1021
      |||||
      DB      388  GCTATCTGAGAAATTAACCTGAGAGCCAGCGAGCTTTCCGAGACATCCGAGACTCT 447

QY      1022  GCTCCCAAGCCCTTCTCATGTGGCGGAGTGGAAAGGAGCATTAAGAGAAATCCCAT 1081
      |||||
      DB      448  GCTCCCAAGCCCTTCTCATGTGGCGGAGTGGAAAGGAGCATTAAGAGAAATCCCAT 507

QY      1082  TGGAAATACAGAAAGAAAGACTCTGTATTAAGAGAGCTGCTCTGAAATTTTCAAAAGTCAA 1141
      |||||
      DB      508  TGGAAATACAGAAAGAAAGACTCTGTATTAAGAGAGCTGCTCTGAAATTTTCAAAAGTCAA 567

QY      1142  AACAAATCTCTTAAGACATCTCTTATTAAGAGAAACCTGTGATCTCTTATCTGAAC 1201
      |||||
      DB      568  AACAAATCTCTTAAGACATCTCTTATTAAGAGAAACCTGTGATCTCTTATCTGAAC 627

QY      1202  TGTTTGTAACGTGAATTGAGAGCCAGAAACCCCAAGAAAGCCCGGTCTTCACTGGA 1261
      |||||
      DB      628  TGTTTGTAACGTGAATTGAGAGCCAGAAACCCCAAGAAAGCCCGGTCTTCACTGGA 687

QY      1262  CACCCACCAAGACTCTCCACTTCCCAAAAGGAGCCAGACCCATCCCAAGAGAGAAAGC 1321
      |||||
      DB      688  CACCCACCAAGACTCTCCACTTCCCAAAAGGAGCCAGACCCATCCCAAGAGAGAAAGC 747

QY      1322  CTTCTGAAGTGAATGAAGATTTGCGGTACCTTTCACACAGAAATGATTTGTGTG 1381
      |||||
      DB      748  CTTCTGAAGTGAATGAAGATTTGCGGTACCTTTCACACAGAAATGATTTGTGTG 807

QY      1382  TTGGCAACAGCCTCCCACTCACCGTTACATTAACGGAATCCTCTCAAGAGAGAG 1441
      |||||
      DB      808  TTGGCAACAGCCTCCCACTCACCGTTACATTAACGGAATCCTCTCAAGAGAGAG 867

QY      1442  GACTGTAGCAGGTGTCTGATGCCATCAAGTGTGAGAGAAATTTCACTTGGCTGT 1501
      |||||
      DB      868  GACTGTAGCAGGTGTCTGATGCCATCAAGTGTGAGAGAAATTTCACTTGGCTGT 927

QY      1502  TCTTTTGGAGGAGCACACTAGTAGAGCTCTTAAGGAGCCCAATCCTTACAGACTTT 1561
      |||||
      DB      928  TCTTTTGGAGGAGCACACTAGTAGAGCTCTTAAGGAGCCCAATCCTTACAGACTTT 987

QY      1562  GGAGAACTGATGACAGTGTGTTTTCGAAGCGGAGTGAAGAACTGAGCTGATGAG 1621
      |||||
      DB      988  GGAGAACTGATGACAGTGTGTTTTCGAAGCGGAGTGAAGAACTGAGCTGATGAG 1047

QY      1622  GAGAGAGAAAGATGGATATTTCAAGATTCAGGGAACAAAGATTTTACAGGACTGCA 1681
      |||||
      DB      1048  GAGAGAGAAAGATGGATATTTCAAGATTCAGGGAACAAAGATTTTACAGGACTGCA 1107

QY      1682  GCTCAAGTATTAAGAAAGAAAGATTCAGGAATCTGAGCCGAGGTTACCTCATTTT 1741
      |||||
      DB      1108  GCTCAAGTATTAAGAAAGAAAGATTCAGGAATCTGAGCCGAGGTTACCTCATTTT 1167

QY      1742  CCTGAGCAGATGATGTTGAAGTTTGAATGATTAACCCCTTCTGCTGTTAGCAT 1801
      |||||
      DB      1168  CCTGAGCAGATGATGTTGAAGTTTGAATGATTAACCCCTTCTGCTGTTAGCAT 1227
```

QY	1802	TGAGACACCATTAACCAAAATTAATCCACAGAAATTTTGAAGCTACCGTGTTGATGAGCG	1861
Dp	1228	TGGACGACCAATTAACCAAAATTAATCTCCACAGAAATTTTGAAGCTACCGTGTTGATGAGCG	1287
QY	1862	TAGCCGATGACGATTTGAGATCCAGAGAAACCAACCTCACCGGACGTTAGGAATA	1921
Dp	1288	TAGCCGATGACGATTTGAGATCCAGAGAAACCAACCTCACCGGACGTTAGGAATA	1347
QY	1922	GCTGTGCTGGCAAGAACCCGTGTCTTCAGATAGTTGTAGCATGCAATTCGCCGAGTGGCA	1981
Dp	1348	GCTGTGCTGGCAAGAACCCGTGTCTTCAGATAGTTGTAGCATGCAATTCGCCGAGTGGCA	1407
QY	1982	GAGACCTGTATATGTGACCTTTGTCTCACTATGTTATCACTGCTGTATATACCTTT	2041
Dp	1408	GAGACCTGTATATGTGACCTTTGTCTCACTATGTTATCACTGCTGTATATACCTTT	1467
QY	2042	CATACCTTCCTGACCTTGTTTTCAATTACTCGATTTTCACAAAACTCTTTCATTCGGCTA	2101
Dp	1468	CATACCTTCCTGACCTTGTTTTCAATTACTCGATTTTCACAAAACTCTTTCATTCGGCTA	1527
QY	2102	ATTGTGAGTTATGAGAGGGTGAATTTGGGATTTCTTTTCCCTTTTGGGAAATGGGCTCTCA	2161
Dp	1528	ATTGTGAGTTATGAGAGGGTGAATTTGGGATTTCTTTTCCCTTTTGGGAAATGGGCTCTCA	1587
QY	2162	AGCTAAAGCTATAGAGATGGCAGATTCAGAGATTCAGGGGTCGTGTTCTATACATTTGCC	2221
Dp	1588	AGCTAAAGCTATAGAGATGGCAGATTCAGAGATTCAGGGGTCGTGTTCTATACATTTGCC	1647
QY	2222	TATGTATAAAGGGGTATAAAGGGCTCTCTTCATTAACATGTGTGAAGATGAAACACCCCTT	2281
Dp	1648	TATGTATAAAGGGGTATAAAGGGCTCTCTTCATTAACATGTGTGAAGATGAAACACCCCTT	1707
QY	2282	CCTTTAAAGCTGTGCTGCTCAGCAGCACTCTTCTCACCCCTGTATACACCCTCTTATATGAGG	2341
Dp	1708	CCTTTAAAGCTGTGCTGCTCAGCAGCACTCTTCTCACCCCTGTATACACCCTCTTATATGAGG	1767
QY	2342	TATATGTATTTTTTAAACCTTAATAATPAAACAAACAACTCACCATGACTTATGAGACGAGA	2401
Dp	1768	TATATGTATTTTTTAAACCTTAATAATAATPAAACAAACAACTCACCATGACTTATGAGACGAGA	1827
QY	2402	AGAGAAATGACAAAGTGAAGCGATGAGAACCCATCTTCAACAGATGAGAAAGACATCGG	2461
Dp	1828	AGAGAAATGACAAAGTGAAGCGATGAGAACCCATCTTCAACAGATGAGAAAGACATCGG	1887
QY	2462	AGAGTGGTATATAACTGTCTGAAAAAGATGTTGTCATTTGAAACATATTCGTGATACA	2521
Dp	1888	AGAGTGGTATATATAACTGTCTGAAAAAGATGTTGTCATTTGAAACATATTCGTGATACA	1947
QY	2522	GTCATGTGGGAAGGGAGTTTGGCTGTGATTAATTTTTCACTATATGATTAACAATTTCT	2581
Dp	1948	GTCATGTGGGAAGGGAGTTTGGCTGTGATTAATTTTTCACTATATGATTAACAATTTCT	2007
QY	2582	TTACTGTCTCAAAAAACCAAAATCTTTGGAAAAAGATGGGAGTGTATGTTTCAGAACAA	2641
Dp	2008	TTACTGTCTCAAAAAACCAAAATCTTTGGAAAAAGATGGGAGTGTATGTTTCAGAACAA	2067
QY	2642	GTTACAGCTGTAAACAAAGACCTTAGATTTTGGGATGGCAGATGCAAAACCTGTATTAAT	2701
Dp	2068	GTTACAGCTGTGTAAACAAAGACCTTAGATTTTGGGATGGCAGATGCAAAACCTGTATTAAT	2127
QY	2702	GTCCTTGTATGACATCACTCTCAAGTATTCCTTCATTTGGGCTTCATCCTTTTATGACAGA	2761
Dp	2128	GTCCTTGTATGACATCACTCTCAAGTATTCCTTCATTTGGGCTTCATCCTTTTATGACAGA	2187
QY	2762	CTCTTGTGTGTGGGATTAAGACCTTAGGAGGGTATGGGGGAGAGTGTGAAATATAGTGTCT	2821
Dp	2188	CTCTTGTGTGTGGGATTAAGACCTTAGGAGGGTATGGGGGAGAGTGTGAAATATAGTGTCT	2247
QY	2822	CCTTTGGCTGGCAAAATGTCTACATCTTTGAAACAAACAGATGTACTTAATGAGCTTCTCCA	2881
Dp	2248	CCTTTGGCTGGCAAAATGTCTACATCTTTGAAACAAACAGATGTACTTAATGAGCTTCTCCA	2307
QY	2882	TTACCTTTGTATAAAAATTAATTTGTATGTATGATGCATCTTGCTCTCCCTCCGCTTTTGT	2941

Db	2308	TTCACTTTGTAATAATATTGTAATGTGTACCAATCTTGGTCTTCCCCCTCCGTTTGT	2367
Qy	2942	TAAATATACAGATAGCACTCCCAAGCCACTTTGGTCTCAGTGTAAATCCCTATTA	3001
Db	2368	TAAATATACAGATAGCACTCCCAAGCCACTTTGGTCTCAGTGTAAATCCCTATTA	2427
Qy	3002	ATCTGAAGAAGAAATAGACCAAGACCTCTGGTCTCAATATATAGAAATTCCTTCTT	3061
Db	2428	ATCTGAAGAAGAAATAGACCAAGACCTCTGGTCTCAATATATAGAAATTCCTTCTT	2487
Qy	3062	TAGCTTCAGGACATATGTGTGTAACAAAGTAGGGGTCTATCTCTAGAAAGTAGGGG	3121
Db	2488	TAGCTTCAGGACATATGTGTGTAACAAAGTAGGGGTCTATCTCTAGAAAGTAGGGG	2547
Qy	3122	TTTTATCCCTTAAAGAAATATGTCCCAAGATATATAGCACTTTTGAAGAACCAAG	3181
Db	2548	TTTTATCCCTTAAAGAAATATGTCCCAAGATATATAGCACTTTTGAAGAACCAAG	2607
Qy	3182	TATGTAGGGTGTGTGGCTGGCCCATCAGTGGACACGAGAGAGATGGATACATTT	3241
Db	2608	TATGTAGGGTGTGTGGCTGGCCCATCAGTGGACACGAGAGAGATGGATACATTT	2667
Qy	3242	GGGAAGAGAAATAAGTTCTCAGGGGCTCCCATCTCTAAGTTTGTGACATGTG	3301
Db	2668	GGGAAGAGAAATAAGTTCTCAGGGGCTCCCATCTCTAAGTTTGTGACATGTG	2727
Qy	3302	ATCTGTGCTTCTGGATTTGACTTTTAAAGAAATTAATCTGGACACATGTATATCT	3361
Db	2728	ATCTGTGCTTCTGGATTTGACTTTTAAAGAAATTAATCTGGACACATGTATATCT	2787
Qy	3362	TGATGATCTTGCTCTTATTTTCTCTTTGTGTGTGTGTGTGTGTGTGTGTGTGT	3421
Db	2788	TGATGATCTTGCTCTTATTTTCTCTTTGTGTGTGTGTGTGTGTGTGTGTGTGT	2847
Qy	3422	GGGTTTCAATTGTGATCTCCATCTGTCTTAGAGAGTGGGCTCTATATAGGAACTGT	3481
Db	2848	GGGTTTCACTGTGATCTCCATCTGTCTTAGAGAGTGGGCTCTATATAGGAACTGT	2907
Qy	3482	GTAACCTCATTTGAGCAAGATGTAGAGAAATPAGACTTAATTCATAGAGGGCTCT	3541
Db	2908	GTAACCTCATTTGAGCAAGATGTAGAGAAATPAGACTTAATTCATAGAGGGCTCT	2967
Qy	3542	CATCTCACACTTAAGAGAGATTTCTAGAAAACTGGGCAAGATTTCTTTGTCTCC	3601
Db	2968	CATCTCACACTTAAGAGAGATTTCTAGAAAACTGGGCAAGATTTCTTTGTCTCC	3027
Qy	3602	ATCATTTTAATGTGGCAGGCTGTTCAGTTTCTTACTCTTAATCTATGTGATTTCTTG	3661
Db	3028	ATCATTTTAATGTGGCAGGCTGTTCAGTTTCTTACTCTTAATCTATGTGATTTCTTG	3087
Qy	3662	TAAAGTGCCAAAAAGAAAAAGAACCCATCAGTGTCTTGACCTTGTCTTTGATCCC	3721
Db	3088	TAAAGTGCCAAAAAGAAAAAGAACCCATCAGTGTCTTGACCTTGTCTTTGATCCC	3147
Qy	3722	TCAGTTTCTTCTGATTTACAGCATGTGTGGGTTCTTAATTTGGATAGATTAGCAAA	3781
Db	3148	TCAGTTTCTTCTGATTTACAGCATGTGTGGGTTCTTAATTTGGATAGATTAGCAAA	3207
Qy	3782	TTTAAACATGTGTGTGTGCCCTAACCAAGGGAATCCCAAGTTTCTGACTTGAAGTAC	3841
Db	3208	TTTAAACATGTGTGTGTGCCCTAACCAAGGGAATCCCAAGTTTCTGACTTGAAGTAC	3267
Qy	3842	TGAGAAGATCACAGAGTGTCTATCTGGCCAGATTTTAAGATGATTTCAATTCCTTG	3901
Db	3268	TGAGAAGATCACAGAGTGTCTATCTGGCCAGATTTTAAGATGATTTCAATTCCTTG	3327
Qy	3902	TCCCTCTCCAGAGACCTCTTATTTTATGTCCCTCTTCTAGTTAATTTCTCTTGA	3961
Db	3328	TCCCTCTCCAGAGACCTCTTATTTTATGTCCCTCTTCTAGTTAATTTCTCTTGA	3387
Qy	3962	TTTGACTTTGTGAGAGAGGTTGACAGTATATAGCAAAAGTTCCAGTGTCAAAATTA	4021

```
Db      3388 TTGACCTTGTGAGAGAGGTTGACAGTAGATTAGCAAAAGTTCGAAAGTGCAAAATTA 3447
Qy      4022 CAGGTGTGTAAGAGTGTGGGGGAAAATTAGCTTATTTTCCCTACATGGAGTACACAC 4081
Db      3448 CAGTGTGTGAGAGTGTGGGGGAAAATTAGCTTATTTTCCCTACATGGAGTACACAC 3507
Qy      4082 TGTGAATTCAGTCTCAAGAGGCGCTGCAAGTCTCCCTAAACATAGTGTGTTGTTT 4141
Db      3508 TGTGAATTCAGTCTCAAGAGGCGCTGCAAGTCTCCCTAAACATAGTGTGTTGTTT 3567
Qy      4142 TCTTTAACAAGTTTAAAGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4201
Db      3568 TCTTTAACAAGTTTAAAGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3627
Qy      4202 CAGCTTTGTTTATAGCCCATTTCAATTTGTTCTGTGTGTTAATTCATTAATCTTTGATA 4261
Db      3628 CAGCTTTGTTTATAGCCCATTTCAATTTGTTCTGTGTGTTAATTCATTAATCTTTGATA 3687
Qy      4262 CCATTTCTGATGTGTAAATGTGTGTTCTGTGTAAATCTTATTAAGAGTTCATTTGTAA 4321
Db      3688 CCATTTCTGATGTGTAAATGTGTGTTCTGTGTAAATCTTATTAAGAGTTCATTTGTAA 3747
Qy      4322 ATAAACTATGTGCTGTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4359
Db      3748 ATAAACTATGTGCTGTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3785
```

```
RESULT 2
BUB57520      927 bp      mRNA      linear      EST 16-OCT-2002
LOCUS      AGENCOURT 10473579 NIH_MGC_107 Homo sapiens cDNA clone
DEFINITION  IMAGE:6647374 5', mRNA sequence.
ACCESSION  BUB57520
VERSION    BUB57520.1 GI:24042512
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 927)
REFERENCE  NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            JOURNAL  Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.jnl.gov
            Plate: LNCM2887 row: k column: 22
            High quality sequence stop: 667.
FEATURES
    source
        1..927
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6647374"
            /rname_type="adenocarcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: breast; Vector: pOTB7; Site 1: EcorI;
            Site 2: XhoI; cDNA made by oligo-dT priming.
            Directionally cloned into EcorI/XhoI sites using the
            following 5' adaptor: GGCACGAG(G). Library constructed by
            Ling Hong in the laboratory of Gerald M. Rubin (University
            of California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH_MGC Library."
```

ORIGIN

```
Query Match      17.7%; Score 776; DB 5; Length 927;
Best Local Similarity 100.0%; Pred. No. 1.6e-291;
Matches 776; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      887 GGTAAAGAAGATTAACGAGAAAGAAAGGACAGAGCTGTTTCAAGGCTATGAACCTGAAGA 946
Db      8 GGTAAAGAAGATTAACGAGAAAGAAAGGACAGAGCTGTTTCAAGGCTATGAACCTGAAGA 67
Qy      947 GAGAGAGAAACAGAGCTATCTGAGAAATTAATCTGAGTCCAGCCGAGCTTTCCGA 1006
Db      68 GAGAGAGAAACAGAGCTATCTGAGAAATTAATCTGAGTCCAGCCGAGCTTTCCGA 127
Qy      1007 GACATCCGAGACTCGCCCTCCGAGGCTTCATGATGGGCGGAGTGAAGGAGATTA 1066
Db      128 GACATCCGAGACTCGCCCTCCGAGGCTTCATGATGGGCGGAGTGAAGGAGATTA 187
Qy      1067 AAGGAAATCCCATTTGGAAGTACAGAAAGAAAGACTCTGTTTAAAGCTGGCTCTGA 1126
Db      188 AAGGAAATCCCATTTGGAAGTACAGAAAGAAAGACTCTGTTTAAAGCTGGCTCTGA 247
Qy      1127 ATTTTCAAAAGTCAAAAACAAAACCTCTAAGCACTCTCTATTAAAGAGAAACCTGTGG 1186
Db      248 ATTTTCAAAAGTCAAAAACAAAACCTCTAAGCACTCTCTATTAAAGAGAAACCTGTGG 307
Qy      1187 TTCTTATCTGAAACCTGTTTGAACGTGAATTTGAGGAGCCAGAAACCCGAAAGGCC 1246
Db      308 TTCTTATCTGAAACCTGTTTGAACGTGAATTTGAGGAGCCAGAAACCCGAAAGGCC 367
Qy      1247 CCGGTCTTCAAGTGAACACCCCAACCAAGACTCTCACTCCCAAAAGGAGCCAGACCCA 1306
Db      368 CCGGTCTTCAAGTGAACACCCCAACCAAGACTCTCACTCCCAAAAGGAGCCAGACCCA 427
Qy      1307 TCCCAAGAGAAAGCCTTCTCAAGTGAATGAAGATTTCCGATCTTTCCACACAGA 1366
Db      428 TCCCAAGAGAAAGCCTTCTCAAGTGAATGAAGATTTCCGATCTTTCCACACAGA 487
Qy      1367 AATGATTTTGTGCTGCTGGACCAAGCTCCCATCACTGATACCAATTCAGGGAATCTTC 1426
Db      488 AATGATTTTGTGCTGCTGGACCAAGCTCCCATCACTGATACCAATTCAGGGAATCTTC 547
Qy      1427 TCCAAAGAAGAGAGAGACTGTAGCAAGGTGTGATGATCAAGTGTGTCAGAGAGAAAC 1486
Db      548 TCCAAAGAAGAGAGAGACTGTAGCAAGGTGTGATGATCAAGTGTGTCAGAGAGAAAC 607
Qy      1487 TTCAGTCTTGGCTGTTCTTTCTTGAAGGACCACTGATGAGCCCTTAAGGAGCCCAAA 1546
Db      608 TTCAGTCTTGGCTGTTCTTTCTTGAAGGACCACTGATGAGCCCTTAAGGAGCCCAAA 667
Qy      1547 TCGTTCAAGACCTTTTGAAGAACCTGATGACAGTGTGTTTGAAGCGGCAATGCAAACT 1606
Db      668 TCGTTCAAGACCTTTTGAAGAACCTGATGACAGTGTGTTTGAAGCGGCAATGCAAACT 727
Qy      1607 GGAGCTGATGAGAGAGAGAAAGAAAGATGGGATATTTCAGAGATCAGGGAACAA 1662
Db      728 GGAGCTGATGAGAGAGAGAAAGAAAGATGGGATATTTCAGAGATCAGGGAACAA 783

RESULT 3
BMS47318      1096 bp      mRNA      linear      EST 20-FEB-2002
LOCUS      AGENCOURT_6507635 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5724505
DEFINITION  5', mRNA sequence.
ACCESSION  BMS47318
VERSION    BMS47318.1 GI:18780990
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 1096)
REFERENCE  NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            JOURNAL  Unpublished (1999)
```

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://lml1.lnl.gov
Plate: LML12713 row: f column: 18
High quality sequence stop: 641.
Location/Qualifiers
1. 1096
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5724305"
/lab_host="DH10B"
/clone_lib="NH_MGC_125"
/note="Organ: ovary (pool of 3); Vector: pCMV-Sport6; Site 1: Scovry (destroyed); Site 2: NotI; RNA source pool of three ovaries, from females ranging in age from 38 to 49 yo. Library is oligo-dT primed and directionally cloned (Scovry site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."

ORIGIN

Query Match 16.9%; Score 741; DB 4; Length 1096;
Best Local Similarity 99.9%; Pred. No. 6.9e-278;
Matches 791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3331 GGAAATATTTGCGACACATGTAATTTCTTGATGATTTGCTGCTTATTTCTCT 3390
DB 21 GGAAATATTTGCGACACATGTAATTTCTTGATGATTTGCTGCTTATTTCTCT 80

QY 3391 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3450
DB 81 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 140

QY 3451 GGAGAGTGGGCTCTCATTAAGGAGACCTGCTGTAACCTTATTCAGCAAGATGTAAG 3510
DB 141 GGAGAGTGGGCTCTCATTAAGGAGACCTGCTGTAACCTTATTCAGCAAGATGTAAG 200

QY 3511 AGAAATAGACTTAATTCATGAGGCTCTCATTCACACCTTAAGAGAGATTTCTA 3570
DB 201 AGAAATAGACTTAATTCATGAGGCTCTCATTCACACCTTAAGAGAGATTTCTA 260

QY 3571 GAAATCTGGGCGAGATTTCTGTTCTCCATCATTTTAATGGAGGCTGTTCACT 3630
DB 261 GAAATCTGGGCGAGATTTCTGTTCTCCATCATTTTAATGGAGGCTGTTCACT 320

QY 3631 TTCTTACTCTTACCTATGATATTTCTTCTGTAACGCTGCCAAAAGAAAAGACCAA 3690
DB 321 TTCTTACTCTTACCTATGATATTTCTTCTGTAACGCTGCCAAAAGAAAAGACCAA 380

QY 3691 TCAGTCTCTTGAATTTGTTCTTGAATCCCTCAATTTCTTGAATTTGAGATGTC 3750
DB 381 TCAGTCTCTTGAATTTGTTCTTGAATCCCTCAATTTCTTGAATTTGAGATGTC 440

QY 3751 GGGTCTCTAATTTGGGATAGATGCAATTAACATTTGTTGTCCTTACCCAG 3810
DB 441 GGGTCTCTAATTTGGGATAGATGCAATTAACATTTGTTGTCCTTACCCAG 500

QY 3811 GGGACTCCCGAGTTTCTGACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3870
DB 501 GGGACTCCCGAGTTTCTGACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 560

QY 3871 CAGATTTAAGTATGATTTCTATTTCTTGGTTCTCTCTCTCTCTCTCTCTCTCTCTCT 3930
DB 561 CAGATTTAAGTATGATTTCTATTTCTTGGTTCTCTCTCTCTCTCTCTCTCTCTCTCT 620

QY 3931 TGTCCTCTCTTCTAGGTAATTTCTCTTGAATTTGACTTTGTTGAAGAGGTTGACA 3990
DB 621 TGTCCTCTCTTCTAGGTAATTTCTCTTGAATTTGACTTTGTTGAAGAGGTTGACA 680

QY 3991 GTGATTTAGGAAGTTCCCAAGTCAAAATTAAGTGTGTTAGTGTGGGGGAAAATTA 4050
DB 681 GTGATTTAGGAAGTTCCCAAGTCAAAATTAAGTGTGTTAGTGTGGGGGAAAATTA 740

QY 4051 GTCTATTTTCTCTCATGAGGATACACCTGTAATTCATCTCACTGAAGGCTT 4110
DB 741 GTCTATTTTCTCTCATGAGGATACACCTGTAATTCATCTCACTGAAGGCTT 800

QY 4111 GCAGTTCTCTCTA 4122
DB 801 GCAGTTCTCTCTA 812

RESULT 4
CN355147 769 bp mRNA linear EST 16-MAY-2004
LOCUS 17000470834451 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
DEFINITION CN355147
ACCESSION CN355147.1 GI:47355081
VERSION CN355147.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 769)
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Muraige, J., Fisk, G. J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R., Lebkowski, J. and Stanton, L. W.
Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 769 Std Error: 0.00.
Location/Qualifiers
1. 769
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, embryoid bodies
derived from H1, H7 and H9 cells"
/clone_lib="GRN_EB"
/note="oligo dT primed, full-length enriched cDNA library from embryoid body outgrowths derived from hES cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions."

ORIGIN

Query Match 16.8%; Score 734; DB 7; Length 769;
Best Local Similarity 100.0%; Pred. No. 3.9e-275;
Matches 734; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1554 GACCTTTGAGAACTCGATGACAGTGTCTTTTCAAGGCGGATGCAAACTGAGCTG 1613
DB 10 GACCTTTGAGAACTCGATGACAGTGTCTTTTCAAGGCGGATGCAAACTGAGCTG 69

QY 1614 GATGAGAGGAAGAAAGATGATTCAGAGATTCAGGGAACAAAGATTTTACAG 1673
DB 70 GATGAGAGGAAGAAAGATGATTCAGAGATTCAGGGAACAAAGATTTTACAG 129

QY 1674 CGACTGAGCTCGAATGTATTAAGAAAGAAATTCAGAAATTCAGCTGAGTTACC 1733
DB 130 CGACTGAGCTCGAATGTATTAAGAAAGAAATTCAGAAATTCAGCTGAGTTACC 189

Db	Query Match	Score	DB	Length	743
4136	TGTTTTCTTAAACAAGTTAAAGCTGTGTATTAATTAATAAATTTG	41.86			
69	TGTTTTCTTAAACAAGTTAAAGCTGTGTATTAATTAATAAATTTG	19			
RESULT 6					
LOCUS	CN355143	743 bp	mRNA	linear	EST 16-MAY-2004
DEFINITION	17000600188342 GRN_PRENEM Homo sapiens CDNA 5', mRNA sequence.				
ACCESSION	CN355143				
VERSION	CN355143.1	GI:47355077			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 743)				
AUTHORS	Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G. J., Li, Y., Xu, C., Pang, R., Guegler, K., Rao, M. S., Mandalam, R., Lebkowski, J. and Stanton, L. W.				
TITLE	Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation				
JOURNAL	Nat. Biotechnol. 22 (6), 707-716 (2004)				
COMMENT	Contact: Brandenberger R Regenerative Medicine Geron Corporation 230 Constitution Drive, Menlo Park, CA 94025, USA Tel: 650 473 8658 Fax: 650 473 7760 Email: rbrandenberger@geron.com Insert Length: 743 Std Error: 0.00. Location/Qualifiers 1..743 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /tissue_type="embryonic stem cell, retinoic acid and mitogen-treated hES cell line H7" /clone_lib="GRN_PRENEM" /note="Oligo dT primed, full-length enriched cDNA library from hES cell line H7 (p29) maintained in feeder-free conditions. Embryoid bodies were generated in the presence of all-trans retinoic acid and mitogens."				
FEATURES					
SOURCE					
ORIGIN					
Query Match	16.1%;	Score 705;	DB 7;	Length 743;	
Best Local Similarity	100.0%;	Pred. No. 7.9e-264;			
Matches 705;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Db	749	GATCCTTGTGTGCAATTTGACCTCATGAAACAGACAGACGCTGCAATTTGAACGTAT	808		
Db	28	GATCCTTGTGTGCAATTTGACCTCATGAAACAGACAGACGCTGCAATTTGAACGTAT	87		
Qy	809	AAAGAGATCGAGAGAGCTGAAGTCAAGAGAGACAGCTCTGCTCGATTTGAACGTAT	868		
Db	88	AAAGAGATCGAGAGAGCTGAAGTCAAGAGAGACAGCTCTGCTCGATTTGAACGTAT	147		
Qy	869	GGAAGGCGGATGCACTGCTTTAAAGAGAGATPACAGAAAGAAAGGCAAGCTGTTTCA	928		
Db	148	GGAAGGCGGATGCACTGCTTTAAAGAGAGATPACAGAAAGAAAGGCAAGCTGTTTCA	207		
Qy	929	GGGCTATGAACTGAGAGAGAGAGAGAAACAGAGCTATCTGAGAAATTTAACTGGAGTG	988		
Db	208	GGGCTATGAACTGAGAGAGAGAGAGAAACAGAGCTATCTGAGAAATTTAACTGGAGTG	267		
Qy	989	CCAGCGGAGCTTTCCGAGACATCCACAGCTCTGCTCCCAAGCCCTTCTCATGTGGCG	1048		
Db	268	CCAGCGGAGCTTTCCGAGACATCCACAGCTCTGCTCCCAAGCCCTTCTCATGTGGCG	327		
Qy	1049	GAGTGGAAAGGACATTAAGAAATTTCCCATTTGGAGTACAGAAAGAAAGACCTCTGT	1108		
Db	328	GAGTGGAAAGGACATTAAGAAATTTCCCATTTGGAGTACAGAAAGAAAGACCTCTGT	387		

FEATURES	source
RESULT 7	
LOCUS	BO014717/c
DEFINITION	BO014717 765 bp mRNA linear EST 26-MAR-2002
ACCESSION	U1-H-BD1-axe-j-04-0-UT.s1 NCI_CGAP_BD1 Homo sapiens cDNA clone
VERSION	IMAGE:5833443 3', mRNA sequence.
KEYWORDS	BO014717
SOURCE	BO014717.1 GI:19739618
ORGANISM	EST.
	Homo sapiens (human)
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE	1 (bases 1 to 765)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
	Tumor Gene Index
	Unpublished (1997)
JOURNAL	Contact: Robert Strausberg, Ph.D.
COMMENT	Email: cgaps-remail.nih.gov
	Tissue Procurement: Dr. Jose Mercuende
	cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
	CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
	Clone Distribution: Clone distribution information can be found
	through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
	Seq primer: M13 FORWARD
	POLYA=yes
	Location/Qualifiers
	1. 765
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon.9606"
	/clone="IMAGE:5833443"
	/tissue_type="Chondrosarcoma"
	/dev_stage="Adult"
	/lab_host="DH10B (Life Technologies)"
	/clone_id="NCI CGAP BD1"
	/note="Organ: Left Pubic Bone; Vector: pT73-Pac
	(Pharmacia) with a modified polylinker; Site 1: EcoR I;
	Site 2: Not I; NCI CGAP BD1 is a normalized cDNA library
	containing the following tissue(s): Chondrosarcoma cell
	line C55. The library was constructed according to
	Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
	1996. First strand cDNA synthesis was primed with an
	oligo-dT primer containing a Not I site. Double stranded
	cDNA was ligated to an EcoR I adaptor, digested with Not
	I, and cloned directionally into pT73-Pac vector. The

oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (d)18 tail. The sequence tag for this library is GCTCAAGGCT.
TAG TISSUE=chondrosarcoma
TAG LIB=UI-H-ED1
TAG_SEQ=GCTCAAGGCT"

ORIGIN

```
Query Match      16.1%; Score 703; DB 5; Length 765;
Best Local Similarity 99.9%; Pred. No. 4.8e-263;
Matches 753; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 946 AAGAGAGAGAAACAGAGCTATCTGAGAAATTAACTGGAGTCCAGCCGGAGCTTTCCG 1005
    |||
Db 765 AAGAGAGAGAAACAGAGCTATCTGAGAAATTAACTGGAGTCCAGCCGGAGCTTTCCG 706

QY 1006 AAGACATCCAGACTGCTGCTCCCAAGCCCTTCTCATGTGGGAGGAGGAGAAAGGACATA 1065
    |||
Db 705 AAGACATCCAGACTGCTGCTCCCAAGCCCTTCTCATGTGGGAGGAGGAGAAAGACATA 646

QY 1066 AAGAGAAATCCCATTTGGAAGTACAGAAAGAAAGAAAGCTGTTAAAGCTGCTCTG 1125
    |||
Db 645 AAGAGAAATCCCATTTGGAAGTACAGAAAGAAAGAAAGCTGTTAAAGCTGCTCTG 586

QY 1126 AATTTTCAAAAGTCAGAAACAAACCTCTTACAGACTCTCTATTAAAGAGAAACCTGTG 1185
    |||
Db 585 AATTTTCAAAAGTCAGAAACAAACCTCTTACAGACTCTCTATTAAAGAGAAACCTGTG 526

QY 1186 GTTCTTATCTGAGAACTGTTTGAACGTGAATTGAGAGCCAGAAAGCCCGAGAAAG 1245
    |||
Db 525 GTTCTTATCTGAGAACTGTTTGAACGTGAATTGAGAGCCAGAAAGCCCGAGAAAG 466

QY 1246 CCGGCTCTTCACTGAGACACCCCAAGACTCTCACTCCCAAGAGAGAGAGAGAGAGAG 1305
    |||
Db 465 CCGGCTCTTCACTGAGAGACACCCCAAGACTCTCACTCCCAAGAGAGAGAGAGAGAGAG 406

QY 1306 ATCCCAAGAGAAAGGCTTCTCAAGTGAATAGAAAGTTTGGCGTACCTTCCAGCAG 1365
    |||
Db 405 ATCCCAAGAGAAAGGCTTCTCAAGTGAATAGAAAGTTTGGCGTACCTTCCAGCAG 346

QY 1366 AATGATATTGTTGTCGTTGGACACAGCTCCCATCAGCGTTACCATTAAGGAAATCC 1425
    |||
Db 345 AATGATATTGTTGTCGTTGGACACAGCTCCCATCAGCGTTACCATTAAGGAAATCC 286

QY 1426 CTCCTCAAGAGAGAGAGAGAGAGAGTGTCTGATGTCATCAAGTGTGTGAGAGAGAA 1485
    |||
Db 285 CTCCTCAAGAGAGAGAGAGAGAGTGTCTGATGTCATCAAGTGTGTGAGAGAGAA 226

QY 1486 CTTCAGTCTTGGCTGTTCTTCTTGGAGGAGACACTCAGTAGAGCTCTAAGGAGAGCCAA 1545
    |||
Db 225 CTTCAGTCTTGGCTGTTCTTCTTGGAGGAGACACTCAGTAGAGCTCTAAGGAGAGCCAA 166

QY 1546 ATCCCTCAGAGCTTTGGAGAACCTGAGTGAAGAGTGTGTTTGGAGGCGGAGAGAGAA 1605
    |||
Db 165 ATCCCTCAGAGCTTTGGAGAACCTGAGTGAAGAGTGTGTTTGGAGGCGGAGAGAA 106

QY 1606 TGGAGCTGATGAGAGAGAGAGAGAGAGTGGAGTATTCAGAGAGTCAAGAGAGAGAA 1665
    |||
Db 105 TGGAGCTGATGAGAGAGAGAGAGAGAGTGGAGTATTCAGAGAGTCAAGAGAGAGAA 46

QY 1666 TTTTACAGGAGCTGAGCTCAGATGTATTAATAA 1699
    |||
Db 45 TTTTACAGGAGCTGAGCTCAGATGTATTAATAA 12

RESULT 8
BQ948768 931 bp mRNA linear EST 21-AUG-2002
LOCUS BQ948768 8805024 Lupski sciatic nerve Homo sapiens cDNA clone
DEFINITION IMAGE:6197047 5', mRNA sequence.
ACCESSION BQ948768
VERSION BQ948768.1 GI:22364246
```

KEYWORDS

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 931)

1 (bases 1 to 931)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-r@mail.nih.gov

Tissue Procurement: Dr. James R. Lupski

cDNA library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLNL13606 row: h column: 08

High quality sequence stop: 640.

Location/Qualifiers

1..931

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6197047"

/sex="male"

/tissue_type="sciatic nerve"

/dev_stage="adult, 70 yr"

/lab_host="DH10B"

/clone_lib="Lupski sciatic nerve"

/note="Vector: pCMV-SPORTS (Life Technologies); Site 1:

NotI; Site 2: SalI; cDNA made by oligo-dT priming.

Directionally cloned using the following adaptors:

5'-TCGACCCAGCGTCCG-3' and

5'-GACTAGTCTGATGCGGCGGCGGCTT(15)-3'. Size selected >

1 kb for average insert length 1.87 kb. This is a primary

library, non-amplified. Library constructed by Life

Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor

College of Medicine) and is available through Life

Technologies."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

Technology."

ORIGIN

```
Query Match      16.0%; Score 698; DB 5; Length 931;
Best Local Similarity 100.0%; Pred. No. 4.1e-261;
Matches 698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3346 GCACATGTAGTATCTTGGATGATCTTGGCTCTTATTCCTTTGTTGTTGTTGTT 3405
    |||
Db 1 GCACATGTAGTATCTTGGATGATCTTGGCTCTTATTCCTTTGTTGTTGTTGTTGTT 60

QY 3406 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3465
    |||
Db 61 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120

QY 3466 TATTAAGGAACTGCTGTAACTTCAATGACAGAGATGTAGAGAAATAGAGACTTAA 3525
    |||
Db 121 TATTAAGGAACTGCTGTAACTTCAATGACAGAGATGTAGAGAAATAGAGACTTAA 180

QY 3526 TTCCATAGAGGCTCATCTCAACCTTAAGAGAGATTTCTAAGAAACCTGGGCGAG 3585
    |||
Db 181 TTCCATAGAGGCTCATCTCAACCTTAAGAGAGATTTCTAAGAAACCTGGGCGAG 240

QY 3586 ATTTCTTTTGTCTCATCATATTTTAAATGTGAGAGCTGTTCAAGTTTCTTAACT 3645
    |||
Db 241 ATTTCTTTTGTCTCATCATATTTTAAATGTGAGAGCTGTTCAAGTTTCTTAACT 300

QY 3646 ATGTGATATTTCTTGTAAAGTGTCCAAAAAGAAAAAGAAAGAAAGAAAGAAAGAAAG 3705
    |||
Db 301 ATGTGATATTTCTTGTAAAGTGTCCAAAAAGAAAAAGAAAGAAAGAAAGAAAGAAAG 360

QY 3706 TTTGTTCTTGTATTCCTCAGTTTCTTCTTGTATTTGAGCATGTGTGGGTTCTTAATTTT 3765
    |||
```


Db		361	TTTGTCTTTATCCCTCAGTTTCCTTGATTACAGATGTCGGTTCCATAATTTG	420				
Oy		3766	GGTATGAGTTAGCAAAATTTAACCATGTGTTTGTCCTACCACGGGGACTCCCAATTT	3825				
Db		421	GGTATGAGTTAGCAAAATTTAACCATGTGTTTGTCCTACCACGGGGACTCCCAATTT	480				
Oy		3826	CTGACTTAAGTAGACTGAGAAGATCCAGCAGGNGCTATCGGCAGATTTAAGTAGAT	3885				
Db		481	CTGACTTAAGTAGACTGAGAAGATCCAGCAGGNGCTATCGGCAGATTTAAGTAGAT	540				
Oy		3886	TCTATTTTCCTTGGTTCCTCCTCTCCCTGAGGACCTGTAATTTAATTTATGCCCTTTAG	3945				
Db		541	TCTATTTTCCTTGGTTCCTCCTCCTCCCTGAGGACCTGTAATTTAATTTATGCCCTTTAG	600				
Oy		3946	GTTAATTTCTCTTTGATTTGACTTTGTTGAGAGAGAGTTGGACAATAGATTACCAAGT	4005				
Db		601	GTTAATTTCTCTTTGATTTGACTTTGTTGAGAGAGAGTTGGACAATAGATTACCAAGT	660				
Oy		4006	TCCAAGTGCAAATTTACAGTGTGTTAGAGTGCTGGGGGG	4043				
Db		661	TCCAAGTGCAAATTTACAGTGTGTTAGAGTGCTGGGGGG	698				
RESULT 9								
Bu682595/c								
LOCUS	Bu682595	697 bp	mRNA	linear EST 07-OCT-2002				
DEFINITION	UI-CF-EC1-act-a-19-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone							
ACCESSION	BU682595							
VERSION	BU682595.1 GI:23533647							
KEYWORDS	EST.							
SOURCE	Homo sapiens (human)							
ORGANISM	Homo sapiens							
REFERENCE AUTHORS TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 697) Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery							
	Genome Res. 6 (9), 791-806 (1996)							
	JOURNAL MEDLINE PUBMED 97044477 8889548							
	COMMENT Contact: McCray, PB McCray Lab University of Iowa 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA Tel: 319 356 4866 Fax: 319 356 7171 Email: paul-mccray@uiowa.edu Tissue Procurement: Dr. M. J. Welsh, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa Clone Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com). Seq primer: M13 FORWARD POLYA=Yes.							
FEATURES								
SOURCE								
Location/Qualifiers								
1..697								
/organism="Homo sapiens"								
/mol_type="mRNA"								
/db_xref="taxon:9606"								
/clone="UI-CF-EC1-act-a-19-0-UI"								
/issue_type="Lung"								
/dev_stage="Adult and Fetal"								
/lab_host="DH10B (Life Technologies) (TI phage resistant)"								
/clone_id="UI-CF-EC1"								
/note="Organ: Lung; Vector: pTTT3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-EC1 is a normalized_cDNA library containing the following tissue(s) : Normal lung from adult and from fetal day 64, day 87, week 19 and week 42. The library was								

constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AAGGCTTAC.

TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 366-371 and 380-383

TAG_LIB=UI -CF-BCI

TAG_SEQ=AAGTCTTAC"

ORIGIN	Query Match	15.9%; Score 697; DB 5; Length 697;
	Best Local Similarity 100.0%; Pred. No. 1,1e-260;	
	Matches 697; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	3662 TAAAGTGTCCAAAAAGAAAAAGAACCCCAATCAGTGTCTCTTGACTTGTGTTTGATGCC	3721
Db	697 TAAAGTGTCCAAAAAGAAAAAGAACCCCAATCAGTGTCTCTTGACTTGTGTTTGATGCC	638
QY	3722 TCAAGTTCTCTTGATTTTCAAGCANGTCTCGGTTCCCAATTTTGGGATGAGTAAAGCAA	3781
Db	637 TCAAGTTCTCTTGATTTTCAAGCANGTGTGGGTTCCCAATTTTGGGATGAGTAAAGCAA	578
QY	3782 TTTTAACCATGTGTGTGTGTGCCCTACCCAGGGGAGCTCCCAAGTTCCTGACTTGAAGTAGAC	3841
Db	577 TTTTAACCATGTGTGTGTGTGCCCTACCCAGGGGAGCTCCCAAGTTCCTGACTTGAAGTAGAC	518
QY	3842 TGAGAGAAATCCACGAGGTGCTATCTGGCCAGATTAAAGTAGATTCTAATTTCTTGTTCC	3901
Db	517 TGAGAGAAATCCACGAGGTGCTATCTGGCCAGATTAAAGTAGATTCTAATTTCTTGTTCC	458
QY	3902 TCCCTCCCTCGAGGACCTCTTAATTTATTTGTCGCCCTCTAGGTAAATTCCTCTTGA	3961
Db	457 TCCCTCCCTCGAGGACCTCTTAATTTATTTGTCGCCCTCTAGGTAAATTCCTCTTGA	398
QY	3962 TTTGACTTTGTGAGAGAGAGTTGGAACAGTAGATTGACAAAGTTCACAGTGCAAAATTA	4021
Db	397 TTTGACTTTGTGAGAGAGAGTTGGAACAGTAGATTGACAAAGTTCACAGTGCAAAATTA	338
QY	4022 CAGTGTGTAGAGTGTGGGGGAAAAATTAAGTCTTAATTTTCCCTACATGGGATACACAC	4081
Db	337 CAGTGTGTAGAGTGTGGGGGAAAAATTAAGTCTTAATTTTCCCTACATGGGATACACAC	278
QY	4082 TGTGAATTCATATCTTCAACTGAAAGGCCCTCAGTCTCTCTAAACATAGTGTGTTGTTT	4141
Db	277 TGTGAATTCATATCTTCAACTGAAAGGCCCTCAGTCTCTCTAAACATAGTGTGTTT	218
QY	4142 TCTTTAACAAGTTTAAGCTAGTGTATAATAATTAAAAAAATTCGTGTCTGTACTT	4201
Db	217 TCTTTAACAAGTTTAAGCTAGTGTATAATAATTAAAAAAATTCGTGTCTGTACTT	158
QY	4202 CAGTTTGTTTTATGCCCATTTCAATATGTGTGTCTGTGTGTATTCATAACTTTTGATA	4261
Db	157 CAGTTTGTTTTATGCCCATTTCAATATGTGTGTGTGTGTATTCATAACTTTTGATA	98
QY	4262 CCAATTTCTGATGTGTAATAATTTGGTGTCTGTGAATAATCTTAATAAGAGTTCAATGTAA	4321
Db	97 CCAATTTCTGATGTGTAATAATTTGGTGTCTGTGAATAATCTTAATAAGAGTTCAATGTAA	38
QY	4322 ATAAACTATTTGGCTGTAAAAAAATTTAAAAAAATTTAAAAAAATTTAAAAAAATTT	4358
Db	37 ATAAACTATTTGGCTGTAAAAAAATTTAAAAAAATTTAAAAAAATTTAAAAAAATTT	1

Accession mRNA sequence.
BI917709
Version BI917709.1 GI:16181507
Keywords EST.
Source Homo sapiens (human)
Organism Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 817)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bms-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1A011625 row: a column: 23
High quality sequence stop: 790.
Location/Qualifiers
1..817
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5247646"
/lab_host="DH10B"
/clone_lib="NIH MGC 121"
/note="Organ: brain; Vector: PCWV-SPORT6; Site_1: NCI; Site_2: EcorV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcorV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."
ORIGIN
Query Match 15.9%; Score 693; DB 4; Length 817;
Best Local Similarity 100.0%; Pred. No. 3.7e-259;
Matches 693; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2597 CAAAATCTTGGAAAAGAGTGGGGATGGTTGTTTACAGACAGTTACAGCTGTAAAC 2656
DB 1 CAAAATCTTGGAAAAGAGTGGGGATGGTTGTTTACAGACAGTTACAGCTGTAAAC 60
QY 2657 AAAAGCATTAGTATTGGATGGCATGCCAAAACCTGTAAATGCTCTGTATACAT 2716
DB 61 AAAAGCATTAGTATTGGATGGCATGCCAAAACCTGTAAATGCTCTGTATACAT 120
QY 2717 CACTTTCAGATATCTCTTCATTTGGCTTATCTTTTACAGAACTCTTGGTGGGA 2776
DB 121 CACTTTCAGATATCTCTTCATTTGGCTTATCTTTTACAGAACTCTTGGTGGGA 180
QY 2777 TAGAGACTTAGAGGAGGTAGGGGAGAGTGTGAAATAGTGGCTCTTGGCTGGCAAA 2836
DB 181 TAGAGACTTAGAGGAGGTAGGGGAGAGTGTGAAATAGTGGCTCTTGGCTGGCAAA 240
QY 2837 TGTCTACATCTTGAACAACAAGATGATCTTAATGAGCTTCTCATTCACCTTTGTAAAA 2896
DB 241 TGTCTACATCTTGAACAACAAGATGATCTTAATGAGCTTCTCATTCACCTTTGTAAAA 300
QY 2897 TAATTGTATGTATACATTTGGTCTCTCTCCCTCCCGTTTGTAAATATCAGATA 2956
DB 301 TAATTGTATGTATACATTTGGTCTCTCTCCCTCCCGTTTGTAAATATCAGATA 360
QY 2957 GACCTCCAGGGCAGCTTGGTCTCAGTGTAAAGTCCCTATTAACTGTAAAGGAAAT 3016

DB 361 GCACTCCAGGCCACTTTGGTCTCAGTGAAGATCCCTATTACTATCTGAAAAGAAAT 420
QY 3017 AGAGCCAAAGACTCTGCTGCTCAATATATAGAAATTCCTTTAGTCTCAGACATA 3076
DB 421 AGAGCCAAAGACTCTGCTGCTCAATATATAGAAATTCCTTTAGTCTCAGACATA 480
QY 3077 TTGTGTGAAAAGAGTGGGGTCTATCTCTCAGAAAGTGGGGCTTTATCTTAAAGA 3136
DB 481 TTGTGTGAAAAGAGTGGGGTCTATCTCTCAGAAAGTGGGGCTTTATCTTAAAGA 540
QY 3137 GAATATGTCCTCCAGATTATTAGCACTTTTGAAGAGAACCAAGATGATGAGGTGTG 3196
DB 541 GAATATGTCCTCCAGATTATTAGCACTTTTGAAGAGAACCAAGATGATGAGGTGTG 600
QY 3197 GCTGGCCCTACGTGAGAGACAGAGAGAAATGGATATACATTGTGGAGAGAGAAAA 3256
DB 601 GCTGGCCCTACGTGAGAGACAGAGAGAAATGGATATACATTGTGGAGAGAGAAAA 660
QY 3257 GTTCTCAGAGGGGCTCCCACTGCTAAAGTTTTT 3289
DB 661 GTTCTCAGAGGGGCTCCCACTGCTAAAGTTTTT 693
RESULT 11
LOCUS CN355144 704 bp mRNA linear EST 16-MAY-2004
DEFINITION 17000532602228 GRN_ES Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN355144
VERSION CN355144.1 GI:47355078
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 704)
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Muraige, J., Fisk, G.J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R., Lebowek, J. and Stanton, L.W.
Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert length: 704 Std Error: 0.00.
Location/Qualifiers
1..704
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, cell lines H1, H7, and H9"
/clone_lib="GRN_ES"
/note="oligo dT primed, full-length enriched cDNA library from undifferentiated hES cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions"
ORIGIN
Query Match 15.8%; Score 689; DB 7; Length 704;
Best Local Similarity 100.0%; Pred. No. 1.4e-257;
Matches 689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1040 ATGTGGCGGAGTGGAAAAGGACATTAAGGAAATCCCATTTGGAGTAGAGAAAGAA 1099
DB 16 ATGTGGCGGAGTGGAAAAGGACATTAAGGAAATCCCATTTGGAGTAGAGAAAGAA 75
QY 1100 GACTCTGTAAAAAGCTGGCTCTGAATTTTCAAAAGTCAAAAAGAAAACTCTAAGCA 1159
DB 76 GACTCTGTAAAAAGCTGGCTCTGAATTTTCAAAAGTCAAAAAGAAAACTCTAAGCA 135

```
QY 1160 CTCTCTATTAAAGAGAACCTGTGTCTCTTAATCTGAAACTGTGTTGTAACGTGAATT 1219
|
|
|
Db 136 CTCTCTATTAAAGAGAACCTGTGTCTCTTAATCTGAAACTGTGTTGTAACGTGAATT 135
|
|
|
QY 1220 GAGAGAGCAAGAAACCCAGAAAGCCCGGTCTTCAGTGAGACACCCACCAAGACTCTC 1279
|
|
|
Db 196 GAGAGAGCAAGAAACCCAGAAAGCCCGGTCTTCAGTGAGACACCCACCAAGACTCTC 255
|
|
|
QY 1280 CACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1339
|
|
|
Db 256 CACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 315
|
|
|
QY 1340 AGATTTGCGCTTACCTTTCACACAGAAATGTAATTTGTGTCGTTGAGACAGCTCCCCC 1399
|
|
|
Db 316 AGATTTGCGCTTACCTTTCACACAGAAATGTAATTTGTGTCGTTGAGACAGCTCCCCC 375
|
|
|
QY 1400 ATCAACCTTACCTTTCACAGAGATCTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAG 1459
|
|
|
Db 376 ATCAACCTTACCTTTCACAGAGATCTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAG 435
|
|
|
QY 1460 GATGCCATCAAGTGTTCAGAGAACTTCACTGTGCTGCTCTTCTTGAAGAGAGAGAG 1519
|
|
|
Db 436 GATGCCATCAAGTGTTCAGAGAACTTCACTGTGCTGCTCTTCTTGAAGAGAGAGAGAG 495
|
|
|
QY 1520 CTCACTAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1579
|
|
|
Db 496 CTCACTAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 555
|
|
|
QY 1580 TGTGTTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1639
|
|
|
Db 556 TGTGTTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 615
|
|
|
QY 1640 TATTTCAGAGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1699
|
|
|
Db 616 TATTTCAGAGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 675
|
|
|
QY 1700 GAAAGAGATTCAGAGATCTGAGCTGAGG 1728
|
|
|
Db 676 GAAAGAGATTCAGAGATCTGAGCTGAGG 704
|
|
|
```

```
RESULT 12
LOCUS BQ229745 870 bp mRNA linear EST 02-MAY-2002
DEFINITION AGENCOURT_7510022 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6042882
5', mRNA sequence.
ACCESSION BQ229745
VERSION BQ229745.1 GI:20411145
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 870)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1AM31282 row: p column: 19
High quality sequence stop: 596.
Location/Qualifiers
1..870
FEATURES
Source
/organism="Homo sapiens"
/mol_type="mRNA"
```

```
/db_xref="taxon:9606"
/clone="IMAGE:6042882"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH MGC 92"
/note="Organ: testis; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; 01igo-dt primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
```

```
Query Match 15.5%; Score 678; DB 5; Length 870;
Best Local Similarity 100.0%; Pred. No. 2.6e-253;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 113 GGGCCCTGCGCGGAGCAATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 172
|
|
|
Db 1 GGGCCCTGCGCGGAGCAATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
|
|
|
QY 173 GGGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 232
|
|
|
Db 61 GGGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
|
|
|
QY 233 CAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 292
|
|
|
Db 121 CAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
|
|
|
QY 293 GGGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 352
|
|
|
Db 181 GGGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
|
|
|
QY 353 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 412
|
|
|
Db 241 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
|
|
|
QY 413 GGGCATTGGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 472
|
|
|
Db 301 GGGCATTGGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
|
|
|
QY 473 GGGCTGCTGCGCATTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 532
|
|
|
Db 361 GGGCTGCTGCGCATTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
|
|
|
QY 533 GGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 592
|
|
|
Db 421 GGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
|
|
|
QY 593 CCCAGCACTACTCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 652
|
|
|
Db 481 CCCAGCACTACTCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
|
|
|
QY 653 GGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 712
|
|
|
Db 541 GGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
|
|
|
QY 713 CTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 772
|
|
|
Db 601 CTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
|
|
|
QY 773 CATGGAACAGACAGACAGCA 790
|
|
|
Db 661 CATGGAACAGACAGACAGCA 678
|
|
|
```

```
RESULT 13
LOCUS BUS17114 888 bp mRNA linear EST 12-SEP-2002
DEFINITION AGENCOURT_10164432 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6514660
5', mRNA sequence.
ACCESSION BUS17114
VERSION BUS17114.1 GI:22824640
KEYWORDS EST.
```



```

QY 3623 GTTCAGTTTCTTACTCTTACTGATGATGATTTCTTGACGTGTCACAAAAAGAAAA 3682
DB 62 GTTCAGTTTCTTACTCTTACTGATGATGATTTCTTGACGTGTCACAAAAAGAAAA 121
QY 3683 AGACCCAAATGAGTCTGACTGTTGTTGATGATGATGATGATGATGATGATGATGAT 3742
DB 122 AGACCCAAATGAGTCTGACTGTTGTTGATGATGATGATGATGATGATGATGATGAT 181
QY 3743 CATGTGCGGGTCTCAATTTTGGGTATGATGATGATGATGATGATGATGATGATGAT 3802
DB 182 CATGTGCGGGTCTCAATTTTGGGTATGATGATGATGATGATGATGATGATGATGAT 241
QY 3803 CTACCCAGGGAGTCTCCAGTCTTCTGATGATGATGATGATGATGATGATGATGAT 3862
DB 242 CTACCCAGGGAGTCTCCAGTCTTCTGATGATGATGATGATGATGATGATGATGAT 301
QY 3863 TATCTGGCCAGATTTAGTATGATTTCTTCTGTTCTCTCTCTCTCTCTCTCTCTCT 3922
DB 302 TATCTGGCCAGATTTAGTATGATTTCTTCTGTTCTCTCTCTCTCTCTCTCTCTCT 361
QY 3923 TATTTATTTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3982
DB 362 TATTTATTTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 421
QY 3983 GTTGCACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4042
DB 422 GTTGCACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 481
QY 4043 GAAATTTAGTCTTATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4102
DB 482 GAAATTTAGTCTTATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 541
QY 4103 AAGGCGCTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4162
DB 542 AAGGCGCTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 601
QY 4163 GTGTATATATATATATATATATATATATATATATATATATATATATATATATAT 4222
DB 602 GTGTATATATATATATATATATATATATATATATATATATATATATATATATAT 661
QY 4223 TCATATTG 4230
DB 662 TCATATTG 669

```

RESULT 15
BM981391/c
LOCUS
DEFINITION
BM981391 715 bp mRNA linear EST 21-FEB-2003
UI-CF-EN1-adh-f-11-0-UI-s1 UI-CF-EN1 Homo sapiens cDNA clone
UI-CF-EN1-adh-f-11-0-UI 3', mRNA sequence.

ACCESSION
BM981391
VERSION
BM981391.1 GI:19603832
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
Normalizaton and subtracction: two approaches to facilitate gene
discovery
JOURNAL
MEDLINE
PUBMED
COMMENT
Genome Res. 6 (9), 791-806 (1996)
97044477
8689548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES
Location/Qualifiers
1. 715
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-adh-f-11-0-UI"
/issue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into pT7T3-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dr)18 tail. The
sequence tag for this library is CTCGTCAGGT.
TAG TISSUE=Human Lung Epithelial Cell lines untreated LPS
6hr to LPS 24h
TAG LIB=UI-CF-EN1
TAG_SEQ=CTGTCAGGT"

ORIGIN
Query Match 15.2%; Score 664; DB 5; Length 715;
Best Local Similarity 99.9%; Pred. No. 7.7e-248;
Matches 714; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 3644 CTATGATATATTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3703
DB 715 CTATGATATATTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 656
QY 3704 ACTTGTCTCTTATCCCTCAGTCTTCTTATGATGATGATGATGATGATGATGATGAT 3763
DB 655 ACTTGTCTCTTATCCCTCAGTCTTCTTATGATGATGATGATGATGATGATGATGAT 596
QY 3764 TGGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3823
DB 595 TGGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 536
QY 3824 TTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3883
DB 535 TTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 476
QY 3884 ATTCTATTTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3943
DB 475 ATTCTATTTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 416
QY 3944 AGGTATATTTCTCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4003
DB 415 AGGTATATTTCTCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 356
QY 4004 GTTCCAAAGTCAAAATTAACAGTGTGATGATGATGATGATGATGATGATGATGATGAT 4063
DB 355 GTTCCAAAGTCAAAATTAACAGTGTGATGATGATGATGATGATGATGATGATGATGAT 296
QY 4064 CTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4123
DB 295 CTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 236

```

QY	4124	AACCTAGTGTGTTGTTTCTTTTCTTTAA	CAAGTTTAA	GCTAGAGTGTAA	TTAAATTTAAAAAAA	4183
Db	235	AACCTAGTGTGTTGTTTCTTTTCTTTAA	CAAGTTTAA	GCTAGAGTGTAA	TTAAATTTAAAAAAA	176
QY	4184	TTGCTGTGTGTCTACTTCA	GCTTGTGTTTAA	AGCCCATTTCA	TATGTGTGTGTGTGT	4243
Db	175	TTGCTGTGTGTCTACTTCA	GCTTGTGTTTAA	AGCCCATTTCA	TATGTGTGTGTGTGT	116
QY	4244	AATTCATTAACCTTTGATACCA	TTTCTGATGTGTAA	AAATTTGGTGTCTGTAA	TATCTTA	4303
Db	115	AATTCATTAACCTTTGATACCA	TTTCTGATGTGTAA	AAATTTGGTGTCTGTAA	TATCTTA	56
QY	4304	TAAAGAGTCAATTTGTAA	ATAA	CTATTTGAGCTTTAA	AAAAAAAAAAAAAAAAAAAA	4358
Db	55	TAAAGAGTCAATTTGTAA	ATAA	CTATTTGAGCTTTAA	AAAAAAAAAAAAAAAAAAAA	1

LOCUS	DEFINITION	EST
CA447821	712 bp mRNA linear	08-NOV-2002
CA447821/c	UT-H-E10-ayf-d-17-0-UI si NCI GAP-E10 Homo sapiens CDNA clone	
CA447821/c	UT-H-E10-ayf-d-17-0-UI 3', mRNA sequence.	

SOURCE	Homo sapiens (human)
...	...

ORGANISM Homo sapiens

REFERENCE	1 (bases 1
1	1

AUTHORS	NCI-CGAP htc
TITLE	National Can

JOURNAL Unpublished
COMMENT Contact: Rob

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaab5-r@email.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@iowa.edu
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES	Loc
source	1.

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cloname="UT-H-E10-ayf-d-17-0-UT"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clonename="NCI CGAP E10"
/notes="Organ: Left Pelvis; Vector: pRT3-Pac (Pharmacia) with a modified polylinker; Site_1: Ecor I; Site_2: Not I. NCI CGAP_E10 is a cDNA library containing the following tissue(s): Chondrosarcoma. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pRT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACACCTTGAC. TAG_TISSUE=Chondrosarcoma TAG_LIB=UT-H-E10 TAG_SEQ=ACACCTTGAC"

```

ORIGIN

Query Match	14.9%	Score 653;	DB 6;	Length 712;
Best Local Similarity	99.9%	Pred. No. 1.5e-243;		
Matches 703; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy	3655	TTCTTCCTAAACGAGTCCAAAAGAAAAAGAACCCAAATCAGAGTCTCTTGACCTTGTCTT	3714
Db	704	TTCTTCGTAAAGTGTCCAAAAGAAAAAGAACCCAAATCAGAGTCTCTTGACCTTGTCTT	645
Qy	3715	TGATCCCTCAAGTTCTTCTTGATTTCAGCATGTGTGGGTCTCTTAATTTGGGATATAGAT	3774
Db	644	TGATCCCTCAAGTTCTTCTTGATTTCAGCATGTGTGGGTCTCTTAATTTGGGATATAGAT	585
Qy	3775	TAGCAAAATTAACCATGTGTGTTGTGACCCTACCCAGGGGAGCTCCCAAGTTCTGACTGGA	3834
Db	584	TAGCAAAATTAACCATGTGTGTTGTGACCCTACCCAGGGGAGCTCCCAAGTTCTGACTGGA	525
Qy	3835	AGTAGACTGAGAAAGATCCACGAGGTGCTATCTGAGCCAGATTTAAGTAGATTTCTATTTCC	3894
Db	524	AGTAGACTGAGAAAGATCCACGAGGTGCTATCTGAGCCAGATTTAAGTAGATTTCTATTTCC	465
Qy	3895	TTGGTTCTCCCTCCCTCCGAGGAGCTGTATTTTATTTATGTGCCCTCTCTAGGTTAAATCT	3955
Db	464	TTGGTTCTCCCTCCCTCCGAGGAGCTGTATTTTATTTATGTGCCCTCTCTAGGTTAAATCT	405
Qy	3955	CCTTTGATTGACTTGTGTGAGAGAGAGTTGACAGTAGATTGACAAAGTTCCAAAGTGC	4014
Db	404	CCTTTGATTGACTTGTGTGAGAGAGAGTTGACAGTAGATTGACAAAGTTCCAAAGTGC	345
Qy	4015	AAAATTACAGTGTGTGAGAGTGGGGGGGAAAATTAGTCTTATTTTCCCTACATGGGAT	4074
Db	344	AAAATTACAGTGTGTGAGAGTGGGGGGGAAAATTAGTCTTATTTTCCCTACATGGGAT	285
Qy	4075	ACAACACTGTGAATTCATCTTCAACTGAGAGGCCCTGCAGTTCCTCTAATAACATAGTGT	4133
Db	284	ACAACACTGTGAATTCATCTTCAACTGAGAGGCCCTGCAGTTCCTCTAATAACATAGTGT	225
Qy	4135	TTGTTTTTCTTTAACAAAGTTTAAAGCTATGTATTAATAATTAATAAAAAAATTCGTGTCTG	4193
Db	224	TTGTTTTTCTTTAACAAAGTTTAAAGCTATGTATTAATAATTAATAAAAAAATTCGTGTCTG	165
Qy	4195	TCTACTTCAGGTTGTTTATGAGCCATTCAATATGTTGTCTGATGTGTAATTCATACT	4253
Db	164	TCTACTTCAGGTTGTTTATGAGCCATTCAATATGTTGTCTGATGTGTAATTCATACT	105
Qy	4255	TTTGATATCCATTTCTGATGTGTAATAATTTGAGTGTCTTGTAATATCTTATATTAAGAGTTCA	4314
Db	104	TTTGATATCCATTTCTGATGTGTAATAATTTGAGTGTCTTGTAATATCTTATATTAAGAGTTCA	45
Qy	4315	ATTGTAAATTAACATATGTGGCTGTAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	4358
Db	44	ATTGTAAATTAACATATGTGGCTGTAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1

RESULT	17
BUI60445	
LOCUS	BUI60445
DEFINITION	BUI60445 921 bp mRNA linear EST 04-SEP-2007
ACCESSION	AGNCNCOURT_7944266 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6149416
VERSION	5, mRNA sequence.
KEYWORDS	BUI60445 BUI60445.1 GI:22674355
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 921) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
AUTHORS	Contact: Robert Strausberg, Ph.D.
TITLE	Email: Gqabds-r@mail.nih.gov
JOURNAL	Tissue Procurement: ATCC/DCTD/DTF
COMMENT	cDNA Library Preparation: Life Technologies, Inc.

Db 402 GTAGCCGATGCGATTTGGAGATTCAGAAAGACAAACCTTCACCGAGCTGTAGAAAT 461
Qy 1921 AGCTGTGTCGGCAAGAACCTGTCTTCAGATAGTTGTAGAGCCATTTCCGAGAGTGC 1980
Db 462 AGCTGTGTCGGCAAGAACCTGTCTTCAGATAGTTGTAGATGATCCATTTCCGAGAGTGC 521
Qy 1981 AAGAGACCTGTATATGTAGACCTTTGTCTTCACATATGTATCACTGCTGTATATACCTT 2040
Db 522 AAGAGACCTGTATATGTAGACCTTTGTCTTCACATATGTATCACTGCTGTATATACCTT 581
Qy 2041 TCATACCTCTTCGACCTTTGTTTCATTTAATCTGATTTTCACAAAACCTTCATTCGGCT 2100
Db 582 TCATACCTCTTCGACCTTTGTTTCATTTAATCTGATTTTCACAAAACCTTCATTCGGCT 641
Qy 2101 AATTGTAGTTATGAGGAGGTGATTTGGATTTCTTTCC 2139
Db 642 AATTGTAGTTATGAGGAGGTGATTTGGATTTCTTTCC 680

RESULT 19
BO898842 891 bp mRNA linear EST 16-AUG-2002
LOCUS AGENCOURT 8121821 lupski dorsal root ganglion Homo sapiens cDNA
DEFINITION BO898842 clone IMAGE:6181536 5', mRNA sequence.
ACCESSION BO898842 GI:22290856
VERSION BO898842.1 GI:22290856
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 891)
NIH-MGC http://mgs.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1AM13566 row: b column: 01
High quality sequence stop: 653.
Location/Qualifiers
1. 891
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6181536"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 YR"
/lab_host="DH10B"
/clone_lib="Lupski dorsal root ganglion"
/note="Vector: PCW-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCGCGTCG-3' and
5'-GACTGATCTTGAATCGGAGCGCGCCCTT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D (Baylor
College of Medicine) and is available through Life
Technologies."

ORIGIN
Query Match 14.6%; Score 637; DB 5; Length 891;
Best Local Similarity 99.7%; Pred. No. 2.5e-237;
Matches 737; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2838 GTCTACATCTTGGAAACAAACAGATGACCTTAATGAGCTTCTGCATTCACCTTTGTAATAAT 2897
Db 6 GTCTACATCTTGGAAACAAACAGATGACCTTAATGAGCTTCTGCATTCACCTTTGTAATAAT 65
Qy 2898 AATTGTATGTATGACATCTTGGTCCCTCCCTCCCTTTGTTAAATATCAGATAG 2957
Db 66 AATTGTATGTATGACATCTTGGTCCCTCCCTCCCTTTGTTAAATATCAGATAG 125
Qy 2958 CACTCCAGGCGCATTTGGTCTCAGTGAAGATCCCTATTAACATCTGAAGAAATA 3017
Db 126 CACTCCAGGCGCATTTGGTCTCAGTGAAGATCCCTATTAACATCTGAAGAAATA 185
Qy 3018 GAGCCAAAGCCTCTGGTCTCAATATATAGAAATTCCTTTCTTATGCTTCAGAGCTAT 3077
Db 186 GAGCCAAAGCCTCTGGTCTCAATATATAGAAATTCCTTTCTTATGCTTCAGAGCTAT 245
Qy 3078 TGTGTAAACAAATGAGGAGGTCTAATCTCTGAAGAGTGGGCTTTATCCCTTAAGAG 3137
Db 246 TGTGTAAACAAATGAGGAGGTCTAATCTCTGAAGAGTGGGCTTTATCCCTTAAGAG 305
Qy 3138 AATATGTCCCAAGATTATTAGACATTTTAGAGAGAGCCAAAGATGTAGAGGTGTGTG 3197
Db 306 AATATGTCCCAAGATTATTAGACATTTTAGAGAGAGCCAAAGATGTAGAGGTGTGTG 365
Qy 3198 CTGGCCCATCAGTGAAGAGAGAGAAATGGATATCATTTGTGGAGAGAGAAAG 3257
Db 366 CTGGCCCATCAGTGAAGAGAGAGAAATGGATATCATTTGTGGAGAGAGAAAG 425
Qy 3258 TTCCCTCAGGGGCTCCCACTGCTAAAGTTTGTGTAGATGTGATCTGTCTCTGGA 3317
Db 426 TTCCCTCAGGGGCTCCCACTGCTAAAGTTTGTGTAGATGTGATCTGTCTCTGGA 485
Qy 3318 TTGACTTTTAAAGAAATATTCTGGCAGACAGATGATTTCTTGATGATCTTCTGC 3377
Db 486 TTGACTTTTAAAGAAATATTCTGGCAGACAGATGATTTCTTGATGATCTTCTGC 545
Qy 3378 TCTTATTTCTCTTTTGT 3437
Db 546 TCTTATTTCTCTTTTGT 605
Qy 3438 CTCATCTGCTTGAAGAGAGTGGGCTCTCTATTAAGGAGACCTGCTGAACCTTCATTCGAG 3497
Db 606 CTCATCTGCTTGAAGAGAGTGGGCTCTCTATTAAGGAGACCTGCTGAACCTTCATTCGAG 665
Qy 3498 CAAGATGTAGAGAGAAATAGACTTAATTCACATGAGGCTCTCATCTCACACTTAAG 3557
Db 666 CAAGATGTAGAGAGAAATAGACTTAATTCACATGAGGCTCTCATCTCACACTTAAG 725
Qy 3558 GAGGAGATTTCTAGAAAA 3576
Db 726 GAGGAGATTTCTAGAAAA 744

RESULT 20
CD652041 789 bp mRNA linear EST 18-JUN-2003
LOCUS AGENCOURT 14553415 NIA Human H1 Embryonic Stem Cell cDNA Library
DEFINITION (long) Homo sapiens cDNA clone IMAGE:30424726 5', mRNA sequence.
ACCESSION CD652041 GI:31888878
VERSION CD652041.1 GI:31888878
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 789)
NIH-MGC http://mgs.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Irene Gins and Mahendra Rao, NIA
 CDNA Library Preparation: Yulan Piao and Minoru Ko
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC
 Clone distribution information
 can be found through the I.M.A.G.E. Consortium (LNL) at:
 http://image.llnl.gov
 Plate: NDAM507 row: m column: 23
 High quality sequence stop: 636.
 Location/Qualifiers

FEATURES

source

1..789
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30424726"
 /tissue_type="Embryonic Stem cells"
 /cell_line="WA01"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_1lb="NIA Human H1 Embryonic Stem Cell cDNA Library (long)"
 /note="Vector: pCMV-Sport6, Site_1: NotI; Site_2: SalI; This is a long-transcript enriched cDNA library (genome Res. 11: 1553-1558 (2001). [PMID: 11544191]) from WA01 cell line. Undifferentiated human ES cell line WA01/H1 was obtained from WiCell Research Institute, Inc., Madison, WI, cultured according to their instructions, on MBF feeders. They formed round colonies with defined edges and were positive for alkaline phosphatase, SSEA-4, OCT3, OCT4, REX1, UTR, TERT, SOX2, CX43 and CX45. They are negative for GATA2, GATA4, PDX1, NCAM, MSX1, SSEA-1, TUBB3, NES, GFAP, and BOMES. When confluent (18-10 days after plating), the ES cells from 4 X 6cm dishes were treated with 1 mg/ml collagenase, type IV (Invitrogen/GIBCO) for 5-10 min and gently scraped off with 5 ml pipette. RNA was purified with Trizol Reagent from Invitrogen. Protocol ref: Genome Res. 11: 1553-1558 (2001). [PMID:11544191] Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-GACTAGTCTGATCGAGCGCGCCGCTTTTCTTTT-3'] from 3.4g of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker lR-SalI, purified by phenol/chloroform extraction, and separated from free linkers by Centricon-100 column. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S for 25 cycles. The products were purified by phenol/chloroform extraction and Centricon-100 column. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-Sport6 plasmid vector. The average insert size is about 3.6kb."

ORIGIN

Query Match 14.5%; Score 632; DB 6; Length 789;
 Best Local Similarity 100.0%; Pred. No. 2.3e-235;
 Matches 632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1453 GGTGTGATGCGCATCAAGTTTCAGAGAACTTCAGTTCGCTTCTTCTTGA 1512
 Db 105 GGTGTGATGCGCATCAAGTTTCAGAGAACTTCAGTTCGCTTCTTCTTGA 164
 QY 1513 GGGACCACTGATGAGCTTAAAGGACCAAACTCTTCAGACCTTTTGGAACTGG 1572
 Db 165 GGGACCACTGATGAGCTTAAAGGACCAAACTCTTCAGACCTTTTGGAACTGG 224
 QY 1573 ATGACAGTGTGTTTTCGAAGCGCATGCAAACTGAGCTGATGAGAAGAGAA 1632
 Db 225 ATGACAGTGTGTTTTCGAAGCGCATGCAAACTGAGCTGATGAGAAGAGAA 284
 QY 1633 GATGGGATATTTCAGAGATCGAGGAACAAAGATTTTACACGCACTGCAGCTGAAATGT 1692

Db 285 GATGGGATATTTCAGAGATCGAGGAACAAAGATTTTACAGCGACTGAGAAATGT 344
 QY 1693 ATAAAAAGAAAGAAATTCAGAAATCTGAGCGCTGAGTTACTCTATTTTCCCTGAGCAG 1752
 Db 345 ATAAAAAGAAAGAAATTCAGAAATCTGAGCGCTGAGTTACTCTATTTTCCCTGAGCAG 404
 QY 1753 ATGATGTTGAAAGTTGATGATTAACCCCTTCTTCTGTTTGATGCAATTTTGGACACCAT 1812
 Db 405 ATGATGTTGAAAGTTGATGATTAACCCCTTCTTCTGTTTGATGCAATTTTGGACACCAT 464
 QY 1813 TACCAAAATTAATCTCCACAGAAATTTTGTAGTACCCCTGTTGGATGAGACCGATGCA 1872
 Db 465 TACCAAAATTAATCTCCACAGAAATTTTGTAGTACCCCTGTTGGATGAGACCGATGCA 524
 QY 1873 GATTGAGATTCAGAAAGCAAAACCTTACCGGAGCTGTAGAAATAGCTGTGCGC 1932
 Db 525 GATTGAGATTCAGAAAGCAAAACCTTACCGGAGCTGTAGAAATAGCTGTGCGC 584
 QY 1933 AAGAACCTGTCTTCAGATGATGTTAGCATGCAATTCGAGAGTGGCAGAGACTGTAT 1992
 Db 585 AAGAACCTGTCTTCAGATGATGTTAGCATGCAATTCGAGAGTGGCAGAGACTGTAT 644
 QY 1993 ATGAGACCTTGTCTTCAGATGATGTTATGCTGCTGATTAATACCTTTTCACTTCTT 2052
 Db 645 ATGAGACCTTGTCTTCAGATGATGTTATGCTGCTGATTAATACCTTTTCACTTCTT 704
 QY 2053 GACTTGTTCATTAATCTGATTTACAA 2084
 Db 705 GACTTGTTCATTAATCTGATTTACAA 736

RESULT 21
 CA307817/c 640 bp mRNA linear EST 05-AUG-2004
 DEFINITION
 UT-H-FT1-bhx-d-08-0-UI.s1 NCI CGAP FT1 Homo sapiens cDNA clone
 CA307817
 CA307817.1 GI:24470871
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITILE
 Tumor Gene Index
 JOURNAL
 Unpublished (1997)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Gary W. Hunningake, U of I
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soaresuiowa.edu
 Seq primer: M13 FORWARD
 PolyA=Yes.

FEATURES

source

Location/Qualifiers
 1..640
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UT-H-FT1-bhx-d-08-0-UI"
 /tissue_type="Alveolar Macrophage"
 /dev_stage="Adult"
 /lab_host="NCI CGAP FT1"
 /note="Organ: Lung; Vector: pRT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: Bcor I; Site_2: Not I;
 NCI CGAP FT1 is a normalized cDNA library constructed from
 a pool of 81 RNA samples from Alveolar Macrophages
 challenged with different treatments. The mRNA samples

we were a mixture of these conditions (times refer to incubations following isolation by bronchoalveolar lavage) (some normal donor macrophages were cultured in some of the conditions, other donor macrophages in different conditions). The mRNA samples were pooled for library construction. Control 0 hours; control 3 hours; control 24 hours; LPS 100 ng/ml, 3 hours; LPS 100 ng/ml, 24 hours; PMA 10 ng/ml, 3 hours; PMA 10 ng/ml, 24 hours; Klebsiella moi 10, 3 hours; Klebsiella moi 10, 24 hours; Staph aureus moi 10, 3 hours; Staph aureus moi 10, 24 hours; Adenoviral vector (Ad5 CMV eGFP), moi 500, 3 hours; Adenoviral vector (Ad5 CMV eGFP), moi 500, 24 hours; wt adenovirus moi 500, 3 hours; wt adenovirus moi 500, 24 hours; Ad vector + LPS 3 hours; Ad vector + LPS 24 hours; wt adenovirus + LPS 3 hours; wt adenovirus + LPS 24 hours. The library was normalized according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GGCCATCGCG. The tissue was provided by Dr. Gary W. Humphreys of the University of Iowa.

TAG_TISSUE=Human Lung Alveolar Macrophage
TAG_LIB=UT-H-PT1
TAG_SEQ=GGCCATCGCG"

Db 402 CCAAGAAAAGGATGAGAGCTGAAGTCAGAGAGACACCGCTCTGCTGGATTG 343
Qy 862 AACGTATGAAAAGGCGATGAGCTGGTAAAGAGATTAACGAGAAAGAGACAGC 921
Db 342 AACGTATGAAAAGGCGATGAGCTGGTAAAGAGATTAACGAGAAAGAGACAGC 283
Qy 922 TGTTCAGGGCTATGAACTGAAAGAGAGAGAGAAAGAGCTATCTGAGAAATTAAC 981
Db 282 TGTTCAGGGCTATGAACTGAAAGAGAGAGAGAAAGAGCTATCTGAGAAATTAAC 223
Qy 982 TGAAGTCCAGCCGAGCTTCCGAGACATCCGAGCTCTGCTCCCAAGCCCTTCAT 1041
Db 222 TGAAGTCCAGCCGAGCTTCCGAGACATCCGAGCTCTGCTCCCAAGCCCTTCAT 163
Qy 1042 GTGGGCGAGTGAAGAGGACATTAAGAGAAATCCCATTTGAGAGTACAGAAAGAA 1101
Db 162 GTGGGCGAGTGAAGAGGACATTAAGAGAAATCCCATTTGAGAGTACAGAAAGAA 103
Qy 1102 CTCCTGTTAAAGAGCTGCTCTGAAATTTTCAAAAGTCAAAACAAAACCTCTAGCACT 1161
Db 102 CTCCTGTTAAAGAGCTGCTCTGAAATTTTCAAAAGTCAAAACAAAACCTCTAGCACT 43
Qy 1162 CTCCTGTTAAAGAGAGAACCTTG 1183
Db 42 CTCCTATTAAGAGAGAACCTTG 21

RESULT 23
BUST17052 902 bp mRNA linear EST 12-SEP-2002
LOCUS AGENCOURT 10164427 NIH_MGC_71 Homo sapiens CDNA clone IMAGE:6514588
DEFINITION 5', mRNA sequence.
BUST17052
VERSION BUST17052.1 GI:22824578
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 902)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strauberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM14090 row: C column: 05
High quality sequence stop: 531.

FEATURES
source
1..902
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6514588"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."

ORIGIN

Query Match 14.2%; Score 621; DB 5; Length 902;
Best Local Similarity 99.9%; Pred. No. 4,46-231;
Matches 671; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2836 ATGTCTACATCTTGAAACAAACAGATGACCTTAATGAGCTTCTTCATTCCTTTGTA 2895
Db 5 ATGTCTACATCTTGAAACAAACAGATGACCTTAATGAGCTTCTTCATTCCTTTGTA 64
Qy 2896 ATAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2955
Db 65 ATAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 124
Qy 2956 AGCACTCCAGGCGACCTTGGTCTCAGTGAATCCCTTAATTAATTAATTAATTAATTA 3015
Db 125 AGCACTCCAGGCGACCTTGGTCTCAGTGAATCCCTTAATTAATTAATTAATTAATTA 184
Qy 3016 TAAGCCAGACCTCTGCTCTCAAAATATATAGAAATGCTTTCTTTAGTCTCAGACT 3075
Db 185 TAAGCCAGACCTCTGCTCTCAAAATATATAGAAATGCTTTCTTTAGTCTCAGACT 244
Qy 3076 ATTGTGAAACAAAGTGAAGGCTTAATCTCTTAAGAGTGAAGGCTTTATCTTTAAG 3135
Db 245 ATTGTGAAACAAAGTGAAGGCTTAATCTCTTAAGAGTGAAGGCTTTATCTTTAAG 304
Qy 3136 AGAATATGCTCCAGATTTATAGACTTTTGAAGAGAGCAAGATGATGATGATGATGAT 3195
Db 305 AGAATATGCTCCAGATTTATAGACTTTTGAAGAGAGCAAGATGATGATGATGATGAT 364
Qy 3196 GAGTGGCCATCAGTGAAGCAGAGAGAGATGATGATGATGATGATGATGATGATGAT 3255
Db 365 GAGTGGCCATCAGTGAAGCAGAGAGAGATGATGATGATGATGATGATGATGATGAT 424
Qy 3256 AGTTCTCAGGGGCTCCCATCTGTAAGTTTGTGAGATGTTGATCTGTCTCTG 3315
Db 425 AGTTCTCAGGGGCTCCCATCTGTAAGTTTGTGAGATGTTGATCTGTCTCTG 484
Qy 3316 GATTGACCTTTTAAAGAAATTTATGAGAGACATGATGATGATGATGATGATGATGAT 3375
Db 485 GATTGACCTTTTAAAGAAATTTATGAGAGACATGATGATGATGATGATGATGATGAT 544
Qy 3376 GCTCTATTTCTCTTTTGT 3435
Db 545 GCTCTATTTCTCTTTTGT 604
Qy 3436 AACTCCATCTGCTTGAAGAGTGGCTCTGATTAAGGAACTGCTTAACCTTCATTC 3495
Db 605 AACTCCATCTGCTTGAAGAGTGGCTCTGATTAAGGAACTGCTTAACCTTCATTC 664
Qy 3496 AGCAGAGATGA 3507
Db 665 AGCAGAGATGA 676

RESULT 24
CK430007 615 bp mRNA linear EST 06-JAN-2004
LOCUS CK430007
DEFINITION Oj43h12.y1 Human lacrimal gland, unamplified: Oj Homo sapiens CDNA
clone Oj43h12.5', mRNA sequence.
CK430007
VERSION CK430007.1 GI:40677668
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 615)
AUTHORS Dickinson, D., Laurie, G. and Wistow, G.
TITLE Expressed sequence tag analysis of human lacrimal gland
JOURNAL Unpublished (2002)
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 43 row: h column: 12

Seq primer: M13R1 reverse primer (ABI).
Location/Qualifiers

1. 615
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="oj43h12"
/issue_type="lacrimal gland"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human lacrimal gland, unamplified: oj"
/note="Organ: Eye; Vector: pCMVSPORT6; RNA was extracted from 2 human lacrimal glands. A directionally cloned cDNA library in the pCMVSPORT6 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of which are contained in the manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor
(5'-pGACTAGTCTAGATCGCAGCGGCCGCTT)15-3'. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Query Match 14.1%; Score 615; DB 7; Length 615;
Best Local Similarity 100.0%; Pred. No. 1e-228; Indels 0; Gaps 0;
Matches 615; Conservative 0; Mismatches 0;

2870 TGAGCTTCCTCCATTCACCTTGTAAATAATTTGTATGTACCATCTTGCCTCCG 2929
1 TGAGCTTCCTCCATTCACCTTGTAAATAATTTGTATGTACCATCTTGCCTCCG 60

2930 CTCCTGTTTTTAAATATCAGATAGCACTCCAGGCCACTTGTCTCAGTGAAGA 2989
61 CTCCTGTTTTTAAATATCAGATAGCACTCCAGGCCACTTGTCTCAGTGAAGA 120

2990 TCCCTATTAATCTATCTGAAGAAGAAATAGAGCCAGCCTTGTCTCAATATATAGA 3049
121 TCCCTATTAATCTATCTGAAGAAGAAATAGAGCCAGCCTTGTCTCAATATATAGA 180

3050 ATGCGCTTCTTATAGTCTTCAAGAGCTATGTGTAAGAAACAAGTGAAGGGGTCTAATCTCTA 3109
181 ATGCGCTTCTTATAGTCTTCAAGAGCTATGTGTAAGAAACAAGTGAAGGGGTCTAATCTCTA 240

3110 GAAAGTATGAGGCTTTATCTTAAAGAAATATGTCCCAATTTATAGCACTTTAGAG 3169
241 GAAAGTATGAGGCTTTATCTTAAAGAAATATGTCCCAATTTATAGCACTTTAGAG 300

3170 GAGAAGCCAAAGTATGAGGCTGTGTGCTGCGCCATCAGTGAAGCAAGAGAGATG 3229
301 GAGAAGCCAAAGTATGAGGCTGTGTGCTGCGCCATCAGTGAAGCAAGAGAGATG 360

3230 GATATCATTTGTGGAAGAGAGAAAGTCTCAGGGGCTCCCACTGTGTAAGTTT 3289
361 GATATCATTTGTGGAAGAGAGAAAGTCTCAGGGGCTCCCACTGTGTAAGTTT 420

3290 TGTGAGATGTGATCTGTGCTTCTGATTTGAATTTAAAGAAATTTCTGGCAGCAC 3349
421 TGTGAGATGTGATCTGTGCTTCTGATTTGAATTTAAAGAAATTTCTGGCAGCAC 480

3350 ATGTAGTATCTTGTGATGATCTGTGCTCTTATTTCTCTTTGTGTGTGTGTGT 3409
481 ATGTAGTATCTTGTGATGATCTGTGCTCTTATTTCTCTTTGTGTGTGTGTGT 540

3410 GTGTGTGGCTATGGGTTTTGATTTGTAACCAATGCTAGAGAGATGGGCTCTCATATA 3469
541 GTGTGTGGCTATGGGTTTTGATTTGTAACCAATGCTAGAGAGATGGGCTCTCATATA 600

3470 AGGGAACCTGCTCTGA 3484
601 AGGGAACCTGCTCTGA 615

RESULT 25
BE896063 640 bp mRNA linear EST 20-OCT-2000
LOCUS 601438632F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3923871 5',
DEFINITION mRNA sequence.
ACCESSION BE896063
VERSION BE896063.1 GI:10360296
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 640)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM9760 row: 1 column: 16
High quality sequence stop: 638.
Location/Qualifiers
1. 640
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3923871"
/issue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/note="Organ: Skin; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

FEATURES

source

ORIGIN

Query Match 14.0%; Score 613; DB 2; Length 640;
Best Local Similarity 100.0%; Pred. No. 6e-228; Indels 0; Gaps 0;
Matches 613; Conservative 0; Mismatches 0;

3445 TGCTTGAAGAGTGGGCTCTATATAGGAACCTGCTGAACCTTATGACAGCAAGAT 3504
1 TGCTTGAAGAGTGGGCTCTATATAGGAACCTGCTGAACCTTATGACAGCAAGAT 60

3505 GTAGAGAAATAGACTTAATTCACCTAGGGCTCTCATCTCACACTTAAGAGAGAG 3564
61 GTAGAGAAATAGACTTAATTCACCTAGGGCTCTCATCTCACACTTAAGAGAGAG 120

3565 TTTCGAAAAAATCGGGCAAGTTTCTTGTCTCCATCATTTATATGTGGCAGCTGT 3624
121 TTTCGAAAAAATCGGGCAAGTTTCTTGTCTCCATCATTTATATGTGGCAGCTGT 180

3625 TCAGTTTCTTACTCTTACCTATGATGATATTTCTTGTAACGCTCAAAAAAGAAAAG 3684
181 TCAGTTTCTTACTCTTACCTATGATGATATTTCTTGTAACGCTCAAAAAAGAAAAG 240

3685 ACCCAATCAGTGTCTTGAATTTGTTGATTCCTCAGATTTCTTGTGATTTAGCA 3744
241 ACCCAATCAGTGTCTTGAATTTGTTGATTCCTCAGATTTCTTGTGATTTAGCA 300

3745 TGTGTGGGTTCTTAATTTGGGTATAGTACAAATTTAACCATTTGTGTGCTCT 3804
301 TGTGTGGGTTCTTAATTTGGGTATAGTACAAATTTAACCATTTGTGTGCTCT 360

3805 ACCCAGGGAGCTCCCAAGTTTCTGACTTGAATGAGCTGAGAGATCCACGAGGTGCTA 3864

QY	2099	CTAATTGTGATGATGAGAGGGATGATTGGGATTTCTTTCCCTTTTGGGAAATGGGCTC	2158
Db	181	CTAATTGTGATGATGAGAGGGATGATTGGGATTTCTTTCCCTTTTGGGAAATGGGCTC	240
QY	2159	TCAAGCTAAAGCTATATAGATGGGAGATTTCAGAAATTTTTCAGGGGCTGTCTTCTATACATTT	2218
Db	241	TCAAGCTAAAGCTATATAGATGGGAGATTTCAGAAATTTTTCAGGGGCTGTCTTCTATACATTT	300
QY	2219	GCCCTATGTTAAAGGGGTAAAGGGGCTCTCTTCAATTAGACATGTGGGAAGATGAAGACACC	2278
Db	301	GCCCTATGTTAAAGGGGTAAAGGGGCTCTCTTCAATTAGACATGTGGGAAGATGAAGACACC	360
QY	2279	CTTCCTTTAGAGCTGTGCTCGACATGGGACCTTTCTCACCTGGTGAACACCTCTTATAGT	2338
Db	361	CTTCCTTTAGAGCTGTGCTCGACATGGGACCTTTCTCACCTGGTGAACACCTCTTATAGT	420
QY	2339	GGGTATATGATATTTTAAACCTTAATAATAAACAACAACCTCACATGAGCTTTAGAAC	2398
Db	421	GGGTATATGATATTTTAAACCTTAATAATAAACAACAACCTCACATGAGCTTTAGAAC	480
QY	2399	AGAAGAGGAATGACCAAGTGAAGGATGAACCAAGCCATCTTCACAGATGAAGAAAGCAT	2458
Db	481	AGAAGAGGAATGACCAAGTGAAGGATGAACCAAGCCATCTTCACAGATGAAGAAAGCAT	540
QY	2459	CGGAGAGTGTGTGATTAACCTGTCTGAAAAGATAGTTGTTCAATTGGAACATTTCTGTGAT	2518
Db	541	CGGAGAGTGTGTGATTAACCTGTCTGAAAAGATAGTTGTTCAATTGGAACATTTCTGTGAT	600
QY	2519	ACAAGTCATGTGGG 2531	
Db	601	ACAAGTCATGTGGG 613	
RESULT 27			
LOCUS	BQ081177	891 bp	mRNA
DEFINITION	AGENCOURT 8065588 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6212597	linear	EST 16-AUG-2002
ACCESSION	BQ081177		
VERSION	BQ081177.1	GI:22273185	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strauberg, Ph.D. Email: cgabs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation cDNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LNCM2376 row: p column: 06 High quality sequence stop: 601. Location/Qualifiers 1..891 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6212597" /tissue_type="epidermoid carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_id="NIH MGC 102" /note="Organ: salivary gland; Vector: pOTB7, Site_1: XhoI; Site_2: EcoRI, cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the		
FEATURES			
source			

following 5' adaptor: GGCAAGAG (G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA Synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 14.0%; Score 611; DB 5; Length 891;
Best Local Similarity 100.0%; Pred. No. 3.5e-227;
Matches 611; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 843 ACCGCTCTGCTGATGTAAGTGAAGGCGATGAGCTGTGAAGAGATTAAC 902
DB 5 ACCGCTCTGCTGATGTAAGTGAAGGCGATGAGCTGTGAAGAGATTAAC 64
QY 903 GAGAAAGAAAGGCAAGCTGTTTCAAGGCTATGAACTGAGAGAGAGAGAG 962
DB 65 GAGAAAGAAAGGCAAGCTGTTTCAAGGCTATGAACTGAGAGAGAGAGAG 124
QY 963 CATCTGAGAAATTTAACTGAGTGCAGCCGAGCTTTCCAGACATCCAGACTTG 1022
DB 125 CATCTGAGAAATTTAACTGAGTGCAGCCGAGCTTTCCAGACATCCAGACTTG 184
QY 1023 CCTCCCAAGCCCTTCTCATGTGGGCGAGTGAAGAGGACATTAATCCCATTT 1082
DB 185 CCTCCCAAGCCCTTCTCATGTGGGCGAGTGAAGAGGACATTAATCCCATTT 244
QY 1083 GGAAGTACAGAAAGAAAGACTCTGTTAAAGCTGGCTCTGAATTTTCAAAAGTCAA 1142
DB 245 GGAAGTACAGAAAGAAAGACTCTGTTAAAGCTGGCTCTGAATTTTCAAAAGTCAA 304
QY 1143 ACAAAATCTCTTAAGACTCTCTATTAAGAGAAACCTGTGTTCTTATCTGAAC 1202
DB 305 ACAAAATCTCTTAAGACTCTCTATTAAGAGAAACCTGTGTTCTTATCTGAAC 364
QY 1203 GTTGTAAAGTGAATTTAGAGAGCAAGAAACCCAGAAAGCCGCTTTCAGTGAG 1262
DB 365 GTTGTAAAGTGAATTTAGAGAGCAAGAAACCCAGAAAGCCGCTTTCAGTGAG 1424
QY 1263 ACCCAAGCAAGACTCTCCCAAAAGAGAGCCAGCAAGCCATCCCAAGAGAGAG 1322
DB 425 ACCCAAGCAAGACTCTCCCAAAAGAGAGCCAGCAAGCCATCCCAAGAGAGAG 484
QY 1323 TTCTCAAGTGAATTTAGAGAGTTCCTTCCACCAAGAAATGTTGTGTCT 1382
DB 485 TTCTCAAGTGAATTTAGAGAGTTCCTTCCACCAAGAAATGTTGTGTCT 544
QY 1383 TGACACAGAGCTCCCATCAGCTTACATTAACGGAATCTCTCAAGAGAGAG 1442
DB 545 TGACACAGAGCTCCCATCAGCTTACATTAACGGAATCTCTCAAGAGAGAGAG 604
QY 1443 ACTGTAGCAAG 1453
DB 605 ACTGTAGCAAG 615

RESULT 28
CB851322 756 bp mRNA linear EST 22-APR-2003
LOCUS UI-CF-ECl-ady-c-18-0-UI-81 UI-CF-ECl Homo sapiens cDNA clone
DEFINITION UI-CF-ECl-ady-c-18-0-UI 3', mRNA sequence.
ACCESSION CB851322.1 GI:30046094
VERSION CB851322
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 756)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477
PUBMED 889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@iowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=No.

FEATURES

source

Location/Qualifiers
1..756
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-ECl-ady-c-18-0-UI"
/tissue_type="lung"
/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-ECl is a normalized cDNA library containing the following tissue(s): Normal lung from adult and from fetal day 64, day 87, week 19 and week 42. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AAGCGTTAC.
TAG_SEQ=None found"

ORIGIN

Query Match 13.9%; Score 607; DB 6; Length 756;
Best Local Similarity 100.0%; Pred. No. 1.3e-225;
Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 782 GCAGCAGCAGCAGCTGCGCCCAAGAAAGAGATCGAGAGCTGAAGTCAGAGAG 841
DB 7 GCAGCAGCAGCAGCTGCGCCCAAGAAAGAGATCGAGAGCTGAAGTCAGAGAG 66
QY 842 CAGGCTCTTGCCTGGATTTGAAGTGAAGAGCGGAGTGCAGCTGTAAAGAGATTA 901
DB 67 CAGGCTCTTGCCTGGATTTGAAGTGAAGAGCGGAGTGCAGCTGTAAAGAGATTA 126
QY 902 CGAGAAAGAAAGGCAAGACTGTTTCAAGGCTATGAACTGAGAGAGAGAGAG 961
DB 127 CGAGAAAGAAAGGCAAGACTGTTTCAAGGCTATGAACTGAGAGAGAGAGAG 186
QY 962 GCTATCTGAGAAATTTAACTGAGAGTGCAGCCGAGCTTTCCAGACATCCAGACTCT 1021
DB 187 GCTATCTGAGAAATTTAACTGAGAGTGCAGCCGAGCTTTCCAGACATCCAGACTCT 246
QY 1022 GCTTCCCAAGCCCTTCTCATGTGGGCGAGTGAAGAGGACATTAAGAGATTTCCCAT 1081
DB 247 GCTTCCCAAGCCCTTCTCATGTGGGCGAGTGAAGAGGACATTAAGAGATTTCCCAT 306
QY 1082 TGGAAGTACAGAAAGAAAGACTCTGTTAAAGAGCTGCTCCGATTTTCAAAAGTCAA 1141
DB 307 TGGAAGTACAGAAAGAAAGACTCTGTTAAAGAGCTGCTCCGATTTTCAAAAGTCAA 366

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.jnl.gov>
 Plate: L1AM9705 row: f column: 19

High quality sequence stop: 691.
 Location/Qualifiers

FEATURES

source

1. 745
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3902610"
 /issue_type="epithelioid carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 70"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1; NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.1 Kb. Library constructed by Life
 Technologies."

ORIGIN

Query Match 13.7%; Score 597; DB 2; Length 745;
 Best Local Similarity 100.0%; Pred. No. 1e-221; Indels 0; Gaps 0;
 Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2780 AGACTTAGGAGGGAGGGAGAGTGTGAAATAGGCTTCCTTGGCTGGCAATGT 2839
 Db 1 AGACTTAGGAGGGAGGGAGAGTGTGAAATAGGCTTCCTTGGCTGGCAATGT 60
 QY 2840 CTACATCTTGAACAACAACAGATGACTTAATGAGCTTCTCATTCACTTTGTAAATAA 2899
 Db 61 CTACATCTTGAACAACAACAGATGACTTAATGAGCTTCTCATTCACTTTGTAAATAA 120
 QY 2900 TTGTATGTGACATCTTGCTCTCCCTCCCTCCGTTTGTAAATAATCAGATAGCA 2959
 Db 121 TTGTATGTGACATCTTGCTCTCCCTCCCTCCGTTTGTAAATAATCAGATAGCA 180
 QY 2960 CTCACAGGCACCTTGGTCTCAGTGTAAAGTCCCTATTAATCTGAAAGAAAATAGA 3019
 Db 181 CTCACAGGCACCTTGGTCTCAGTGTAAAGTCCCTATTAATCTGAAAGAAAATAGA 240
 QY 3020 GCCAAGACCTCTGCTCAAAATATATAGAAATGCTTCTTATAGTCTCAGACTATTTG 3079
 Db 241 GCCAAGACCTCTGCTCAAAATATATAGAAATGCTTCTTATAGTCTCAGACTATTTG 300
 QY 3080 TGTGAACAAGTAGGGGTCTATCTCCTAGAAAGTAGGGGCTTTATCCTTAAAGAA 3139
 Db 301 TGTGAACAAGTAGGGGTCTATCTCCTAGAAAGTAGGGGCTTTATCCTTAAAGAA 360
 QY 3140 TATGTCCCAAGATTATTAAGCACTTTTAAAGAGAACCAAGTATGAGGTGTGCTGCT 3199
 Db 361 TATGTCCCAAGATTATTAAGCACTTTTAAAGAGAACCAAGTATGAGGTGTGCTGCT 420
 QY 3200 GGCCCATCATGTGAGACAGAAAGAGAAATGCAATCTTGTGGAAAGAGAAAGAAAGTT 3259
 Db 421 GGCCCATCATGTGAGACAGAAAGAGAAATGCAATCTTGTGGAAAGAGAAAGAAAGTT 480
 QY 3260 CCTCAGGGGCTCCCACTGCTAAAGTTTTTGTGAAGTGTGATCTGTGCTTCTTGATTT 3319
 Db 481 CCTCAGGGGCTCCCACTGCTAAAGTTTTTGTGAAGTGTGATCTGTGCTTCTTGATTT 540
 QY 3320 TGACTTTTAAAGAAATTAATCTGAGACACATGATGATTTCTGGATGATCTTGCTG 3376
 Db 541 TGACTTTTAAAGAAATTAATCTGAGACACATGATGATTTCTGGATGATCTTGCTG 597

RESULT 31
 BE693636 631 bp mRNA linear EST 11-SEP-2000
 LOCUS BE693636
 DEFINITION MR4-BT0358-120600-009-c01 BT0358 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE693636

VERSION BE693636.1 GI:10080796

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

CONTACT

LABORATORY

RUA

PROF.

ANTONIO

PRUDENTE

109,

4

ANDAR,

01509-010,

SAO PAULO-SP,

BRAZIL

TEL: +55-11-2704922

FAX: +55-11-2707001

EMAIL: asimpson@ludwig.org.br

THIS

SEQUENCE

WAS

DERIVED

FROM

THE

FAPESP/LICR

HUMAN

CANCER

GENOME

PROJECT.

THIS

ENTRY

CAN

BE

FOUND

IN

THE

FOLLOWING

URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR4-BT0358-120600-009-c01&cl=2>)

SEG

PRIMER:

PUC

18

FORWARD

HIGH

QUALITY

SEQUENCE

START:

17

HIGH

QUALITY

SEQUENCE

STOP:

606.

LOCATION

SEQUENCE

QUALIFIERS

1. 631

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="BT0358"

/note="Organ: breast; Vector: puc18; Site: 1; Smal; Site: 2; Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 13.6%; Score 594; DB 2; Length 631;
 Best Local Similarity 100.0%; Pred. No. 1.6e-220; Indels 0; Gaps 0;
 Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 860 TGAACCTATGAAGAACCGGATGACCTGTTAAAGAGTAACGAAAGAAAGGACAA 919
 Db 22 TGAACCTATGAAGAACCGGATGACCTGTTAAAGAGTAACGAAAGAAAGGACAA 81
 QY 920 GCTGTTTCAAGGCTATGAAGTAAGAGAGAGAGAAACAGACTTGTCTGAAGAAATTTAA 979
 Db 82 GCTGTTTCAAGGCTATGAAGTAAGAGAGAGAGAAACAGACTTGTCTGAAGAAATTTAA 141
 QY 980 ACTGAGTCCAGCCGAGACTTCCGAGACATCCAGACTTGCCTCCAGCCCTTCTC 1039
 Db 142 ACTGAGTCCAGCCGAGACTTCCGAGACATCCAGACTTGCCTCCAGCCCTTCTC 201
 QY 1040 ATGTGGCGGAGTGAAGAGACATTAAGAGAAATCCCATTTGGAAAGTACAGAAAGAA 1099
 Db 202 ATGTGGCGGAGTGAAGAGACATTAAGAGAAATCCCATTTGGAAAGTACAGAAAGAA 261
 QY 1100 GACTCTGTTAAAGACCTGCTCCGTAATTTTCAAAAGTCAAAACAAAGCTCTTAAGCA 1159
 Db 262 GACTCTGTTAAAGACCTGCTCCGTAATTTTCAAAAGTCAAAACAAAGCTCTTAAGCA 321


```
QY 1160 CTCTCTATTAAAGAGAACCTGTGTTCTTATCTGAAACTGTTTGTAAAGTGAATT 1219
Db 322 CTCTCTATTAAAGAGAACCTGTGTTCTTATCTGAAACTGTTTGTAAAGTGAATT 381
QY 1220 GAGGAGCAAGAAACCCAGAAAAGCCGGTCTTCAGTGAGACACCCACCAAGACTCTC 1279
Db 382 GAGGAGCAAGAAACCCAGAAAAGCCGGTCTTCAGTGAGACACCCACCAAGACTCTC 441
QY 1280 CACTCCCAAAAGGAGACCCAGACACCCATCCCAAGAGAAAAGCTTCAAGTGAATAGA 1339
Db 442 CACTCCCAAAAGGAGACCCAGACACCCATCCCAAGAGAAAAGCTTCAAGTGAATAGA 501
QY 1340 AGATTGGCGTACCTTCCACCAAGATGATTTGTGTGTTGSCACCAAGCTCCCC 1399
Db 502 AGATTGGCGTACCTTCCACCAAGATGATTTGTGTGTTGSCACCAAGCTCCCC 561
QY 1400 ATACCGTTACCTTACCGGAATCTCTCCAAAGAGAGAGACCTGTACCAAG 1453
Db 562 ATACCGTTACCTTACCGGAATCTCTCCAAAGAGAGAGACCTGTACCAAG 615

RESULT 32
BX952313 644 bp mRNA linear EST 01-MAR-2004
LOCUS DKFZ781M12183.r1 781 (synonym: hicc4) Homo sapiens cDNA clone
DEFINITION DKFZ781M12183 5', mRNA sequence.
ACCESSION BX952313
VERSION BX952313.1 GI:43428943
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 644)
Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Mleemann,S.
EST (Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., et al.)
Unpublished (2003)
CONTACT: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp781M12183) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Source
1..644
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp781M12183"
/dev_stage="adult"
/lab_host="DH108"
/clone_lib="781 (synonym: hicc4)"
/notes="Vector: pSport1_Sfi; Site_1: SfiI; Site_2: SfiIB;
cDNA-collection"

ORIGIN
Query Match 13.6%; Score 593; DB 5; Length 644;
Best Local Similarity 99.8%; Pred. No. 3.8e-220;
Matches 643; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 785 GCAGCAGCAGCTGAGGCGCAAGAGAGAGATGAGAGCTGAAGTCAGAGAGAGAC 844
Db 1 GCAGCAGCAGCTGAGGCGCAAGAGAGAGATGAGAGCTGAAGTCAGAGAGAGAC 60
QY 845 GCTCTCTTCTCGATTGATGAGTAAAGCGGATCAGCTGTAAAGAGATTAACGA 904
|||||
```

```
Db 61 GCTCTCTTCTCGGATTGAAAGCTATGAAAGGCGGATGCACTGTGTAAGAGATTAACGA 120
QY 905 GAAAGAAAGCACAAGCTGTTTCAGGCTATGAACTGAAGAGAGAGAAA CAGAGCT 964
Db 121 GAAAGAAAGCACAAGCTGTTTCAGGCTATGAACTGAAGAGAGAGAAA CAGAGCT 180
QY 965 ATCTGAGAAATTAATTAAGTGAAGTCCAGGCGGAGCTTTCCGAGACATCCAGACTCTGCC 1024
Db 181 ATCTGAGAAATTAATTAAGTGAAGTCCAGGCGGAGCTTTCCGAGACATCCAGACTCTGCC 240
QY 1025 TCCCAAGCCCTTCTCATGTGGGGGAGTGGAGAAAGGACATTAAGAGAAATCCCATTTGG 1084
Db 241 TCCCAAGCCCTTCTCATGTGGGGGAGTGGAGAAAGGACATTAAGAGAAATCCCATTTGG 300
QY 1085 AAGTACGAAAGAAAGACTCTCTGTTAAAGAGTGGCTCTGAAATTTTCAAAAGTCAAAAC 1144
Db 301 AAGTACGAAAGAAAGACTCTCTGTTAAAGAGTGGCTCTGAAATTTTCAAAAGTCAAAAC 360
QY 1145 AAAAATCTTACGACTCTCTCTATTAAAGAGAACCTGTGTTCTTATCTGAAACTGT 1204
Db 361 AAAAATCTTACGACTCTCTCTATTAAAGAGAACCTGTGTTCTTATCTGAAACTGT 420
QY 1205 TTGTAAACGTGATTTGAGAGGCCAAGAAACCCAGAAAGCCCGGTCTTCAATGAGAC 1264
Db 421 TTGTAAACGTGATTTGAGAGGCCAAGAAACCCAGAAAGCCCGGTCTTCAATGAGAC 480
QY 1265 CCCACCAAGACTCTCCACTCCCAAAAGGAGCCAGACCCATCCCAAGAGAGAAAGCCTT 1324
Db 481 CCCACCAAGACTCTCCACTCCCAAAAGGAGCCAGACCCATCCCAAGAGAGAAAGCCTT 540
QY 1325 CTCAGTGAGATGAGATTTGGCGTACCTTTCCACCAAGAAATGATTTGTGCTTG 1384
Db 541 CTCAGTGAGATGAGATTTGGCGTACCTTTCCACCAAGAAATGATTTGTGCTTG 600
QY 1385 GCACCAAGCTCTCCCAATCACCTTACCATTAAGGGAATCTCTTC 1428
Db 601 GCACCAAGCTCTCCCAATCACCTTACCATTAAGGGAATCTCTTC 644
```

```
RESULT 33
CD639095 716 bp mRNA linear EST 17-JUN-2003
LOCUS AGENCOURT 14532950 NIH MGC 191 Homo sapiens cDNA clone
DEFINITION IMAGE:30418364 5', mRNA sequence.
ACCESSION CD639095
VERSION CD639095.1 GI:31805567
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 716)
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
CONTACT: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics / NIH
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Narayan Bhat
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
Plate: NDCM21 row: d column: 21
High quality sequence stop: 569.
FEATURES
Source
1..716
/organism="Homo sapiens"
/mol_type="mRNA"
```

/db_xref="taxon:9606"
 /clone="IMAGE:30418364"
 /tissue_type="Pooled"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_1ib="NIH_MGC_191"
 /note="Vector: pDNR-LIB; Site 1: SfiI (ggccatcagcc);
 Site 2: SfiI (ggccgcctccgc); Library is oligo-dT primed
 and directionally cloned. PBMC: Peripheral Blood
 Mononuclear Cells. RNA was pooled from 3/6hour stimulation
 with PMA adn Ionomycin. 5' and 3' adaptors were used in
 cloning as follows: 5' adaptor sequence:
 5'-ATTCTAGAGCGCGAGCGCCGACAG-dt(30)-BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.69
 kb (range 0.70-5.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC library."

ORIGIN

Query Match 13.4%; Score 587; DB 6; Length 716;
 Best Local Similarity 100.0%; Pred. No. 8.2e-218;
 Matches 587; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1305 CATCCCAAGAGAAAGCTTCTCAAGTGAATGAAGATTGGCCATCTTCACACACA 1364
 3 CATCCCAAGAGAAAGCTTCTCAAGTGAATGAAGATTGGCCATCTTCACACACA 62

1365 GAATGTATTTGTGTGGTGGACCAAGCTCCCTCCATCACTTACCATTCAGGATCC 1424
 63 GAATGTATTTGTGTGGTGGACCAAGCTCCCTCCATCACTTACCATTCAGGATCC 122

1425 TCTCCAAAGAGAGAGACTGTAGCAAGGTGTCTGATGCCATCAAGTTTGCAGAGAA 1484
 123 TCTCCAAAGAGAGAGACTGTAGCAAGGTGTCTGATGCCATCAAGTTTGCAGAGAA 182

1485 ACTTCAGTGTGGCTTCTCTTTGAGAGGACACCTAGTAGAGCTCTTAAGAGGACCA 1544
 183 ACTTCAGTGTGGCTTCTCTTTGAGAGGACACCTAGTAGAGCTCTTAAGAGGACCA 242

1545 AATCCTTCAGACCTTTTGGAGAACTGATGACAGTGTGTTTTCGAAGCGCATGCAAAA 1604
 243 AATCCTTCAGACCTTTTGGAGAACTGATGACAGTGTGTTTTCGAAGCGCATGCAAAA 302

1605 CTGGAGCTGATGAGAGAGAGAAAGATGGATATTCAGAGATCAGAGAAACAAG 1664
 303 CTGGAGCTGATGAGAGAGAGAAAGATGGATATTCAGAGATCAGAGAAACAAG 362

1665 ATTTTACAGCAGCTGACGCTCAGATGTATTAATAAAGAAAGATTCAGAGATTCAGGCT 1724
 363 ATTTTACAGCAGCTGACGCTCAGATGTATTAATAAAGAAAGATTCAGAGATTCAGGCT 422

1725 GAGGTACCTCATTTTCCCTGAGCAGATGATGTTGAAGTTGATGATTAACCCCTTC 1784
 423 GAGGTACCTCATTTTCCCTGAGCAGATGATGTTGAAGTTGATGATTAACCCCTTC 482

1785 TTGCTGTGTGATGATTTGGACGACCATTAACAAATTAATCTCAGAGATTTTGAAGCTA 1844
 483 TTGCTGTGTGATGATTTGGACGACCATTAACAAATTAATCTCAGAGATTTTGAAGCTA 542

1845 CCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1891
 543 CCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 589

RESULT 34
 BU781747 593 bp mRNA linear EST 11-OCT-2002
 LOCUS BU781747 1m94e12.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:6122614
 DEFINITION 5', mRNA sequence.
 ACCESSION BU781747
 VERSION BU781747.1 GI:23824324
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 593)
 AUTHORS Lemshka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Blistein, A.,
 Schmitt, A., Theising, B., Ritzer, E., Ronko, I., Bennett, J.,
 Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarashvili, R.,
 Williams, T., Jackson, Y., and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)

TITLE JOURNAL
 COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138

Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. J. Ferrer in vivo mass-excised to
 pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
 University Genome Sequencing Center for information on obtaining a
 clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
 Seq primer: 40RP from Gibco
 High quality sequence stop: 430.
 Location/Qualifiers

FEATURES

1..593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6122614"
 /tissue_type="insulinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="Human insulinoma"
 /note="Organ: pancreas; Vector: pBluescript SK-; Site 1:
 XhoI; Site 2: EcoRI; Constructed with lambda ZAP1 system
 (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
 pBluescript SK- by Dr. H. Inoue following the Washington
 University protocol
 (http://genome.wustl.edu/esclambda_protocol.shtml).
 Please contact Hiroshi Inoue, MD/PhD for further
 information on this library (Metabolism Division, Permutt
 Laboratory, Washington University School of Medicine, Box
 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
 is a Washington University Pancreas EST project library."

ORIGIN

Query Match 13.4%; Score 586; DB 5; Length 593;
 Best Local Similarity 100.0%; Pred. No. 2.1e-217;
 Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

634 CGCGCAACCGCGGAGACCTGGCGGCGAGGAGGAGATGAGAGAGAGAGCC 693
 8 CGCGCAACCGCGGAGACCTGGCGGCGAGGAGGAGATGAGAGAGAGAGCC 67

694 CTCTCGGGGGTGTGGCGGCTCGGAGCTTCCAGTGAAGCCGCTGCTCAACAGATCC 753
 68 CTCTCGGGGGTGTGGCGGCTCGGAGCTTCCAGTGAAGCCGCTGCTCAACAGATCC 127

754 TTCTGCTGCAATTTGACCTCATCGAACACAGACGACGCTGCGGCGCAAGAAAG 813
 128 TTCTGCTGCAATTTGACCTCATCGAACACAGACGACGCTGCGGCGCAAGAAAG 187

814 AGATTCAGAGAGCTGAGTCAAGAGAGAGACGCTCTCTGCTGATGAGATGAGAA 873
 188 AGATTCAGAGAGCTGAGTCAAGAGAGAGACGCTCTCTGCTGATGAGATGAGAA 247

874 GCGGATGACAGCTGTAAAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 933
 248 GCGGATGACAGCTGTAAAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 307

QY	934	ATGAAACGGAAGAGAGAGAAAACAGAGCTATCTGAGAAAAATTAACTGGAGTGCCAGC	993
Db	308	ATGAAACTGAAAGAGAGAGAAACAGAGCTATCTGAGAAAAATTAACTGGAGTGCCAGC	367
QY	994	CGAGACTTCCGAGACATCCGAGACTCTGCGCTCCCAAGCCCTTCTCATGTGCGGAGGTG	1053
Db	368	CGAGACTTCCGAGACATCCGAGACTCTGCGCTCCCAAGCCCTTCTCATGTGCGGAGGTG	427
QY	1054	GAAGGGAACATPAAAAGAAATCCCATTTGGAGTACAGAAAGAAAGACTCTGTTAAA	1113
Db	428	GAAGGGAACATPAAAAGAAATCCCATTTGGAGTACAGAAAGAAAGACTCTGTTAAA	487
QY	1114	AGCTGGCCCTCGAATTTTCAAAAGTCAAAACAAAACCTTAAGCATCTCCTATTAAAG	1173
Db	488	AGCTGGCCCTCGAATTTTCAAAAGTCAAAACAAAACCTTAAGCATCTCCTATTAAAG	547
QY	1174	AGGAACCTCTGTGTTCTTATCTGAAACTGTTTGTAAACGTGAATT	1219
Db	548	AGGAACCTCTGTGTTCTTATCTGAAACTGTTTGTAAACGTGAATT	593
RESULT 35			
BI820795/c			
LOCUS	603034002p1 NIH_MGC_115	Homo sapiens	cdNA clone IMAGE:5175097 5',
DEFINITION			mRNA sequence.
ACCESSION	BI820795		
VERSION	BI820795.1	GI:15932345	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 697)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: LLM411436 row: c column: 02 High quality sequence stop:693.		

```

FEATURES
SOURCE
    location/Qualifiers
    1. .697
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:5175097"
    /lab_host="DH10B"
    /clone_id="NH MGC 115"
    /note="Organ: pooled brain, lung, testis; Vector:
    pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
    source anonymous pool of 6 male brains, age range 23-27; 1
    male lung, age 27; and 1 male testis, age 69. Library is
    oligo-dT primed and directionally cloned (EcoRV site is
    destroyed upon cloning). Average insert size 1.8 kb,
    insert size range 1-3 kb. Library is normalized and
    enriched for full-length clones and was constructed by C.
    Gruber (Invitrogen). Research Genetics tracking code
    021. Note: this is a NIH MGC Library."

```

ORIGIN	Query Match	13.2%	Score 576;	DB 4;	Length 697;
	Best Local Similarity	99.9%	Pred No.1.6e-213;		
	Matches 696;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1.
OY	2103	TTGTGAGTTTANGAGAGGTGATTTGGATTCTTTTCCCTTTTGGGAANAAGGACTTC	CA	2162	

Db	697	TTTGAGATTAGAGGGGATTGGGATTTCTTTTCCCTTTTGGGAAATGGGCTCCAA	638
Qy	2163	GCTAAGCTATAGAGTGGCAGATTCAGAAAGTTTCAGGGGTCTGTTCTATACATTTGGCT	2222
Db	637	GCTAAGCTATAGATGGCAGATTCAAGATTTTCAGGGGTCTGTTCTATACATTTGGCT	578
Qy	2223	ATGTTAAAGGGGTAAAGGGCTCTCTTCATTAGA CATTGGAGAGATGAAGAGCCCTTC	2282
Db	577	ATGTTAAAGGGGTAAAGGGCTCTCTTCATTAGA CATTGGAGAGATGAAGAGCCCTTC	518
Qy	2283	CTTTAGAGCTGTGCTGCATGAGCACTCTTCA CCGTGGTACCCCTCTATATGGGGT	2342
Db	517	CTTTAGAGCTGTGCTGCATGAGCACTCTTCA CCGTGGTACCCCTCTATATGGGGT	458
Qy	2343	ATAGTGAATTTTAAACCTTA-AAATAAACAAACCAACCTCACATGACCTTTAGACCA	2401
Db	457	ATAGTGAATTTTAAACCTTAAGAAATAAACAAACCAACCTCACATGACCTTTAGACCA	398
Qy	2402	AGAGAGATGACAAATGAAAGCATACACAAAGCCATCTTCACAGTGAAGAAAGCATCG	2461
Db	397	AGAGAGATGACAAATGAAAGCATACACAAAGCCATCTTCACAGTGAAGAAAGCATCG	338
Qy	2462	AGAGTGGTGAATTAAGTCTGTGAAGAAATAGTTGTTCA TTGGAACCTATCTGTGATCA	2521
Db	337	AGAGTGGTGAATTAAGTCTGTGAAGAAATAGTTGTTCA TTGGAACCTATCTGTGATCA	278
Qy	2522	GTCATGTGGGAGGGATGTTGGCTGTGATTA TTTTTCAGTAAATGATTAACAATTTCT	2581
Db	277	GTCATGTGGGAGGGATGTTGGCTGTGATTA TTTTTCAGTAAATGATTAACAATTTCT	218
Qy	2582	TTACTGCTCAAAACCAAATCTTTGGAAAAAGAGGGGATGCTTAAGTTTCGAACAA	2641
Db	217	TTACTGCTCAAAACCAAATCTTTGGAAAAAGAGGGGATGCTTAAGTTTCGAACAA	158
Qy	2642	GTTACAGCTGAACAAAGACCTTAGATTTGGGATGGCATGCGCAAACTGTATAAT	2701
Db	157	GTTACAGCTGAACAAAGACCTTAGATTTGGGATGGCATGCGCAAACTGTATAAT	98
Qy	2702	GTCCTTGATCACATCACTTCTCAAGTAATTC TTA TTGGGCTTTCATCTTTTAGCAGAA	2761
Db	97	GTCCTTGATCACATCACTTCTCAAGTAATTC TTA TTGGGCTTTCATCTTTTAGCAGAA	38
Qy	2762	CTCTTGATGATGAGATAGAGCTTAAGGAGGATGGG	2798
Db	37	CTCTTGATGATGAGATAGAGCTTAAGGAGGATGGG	1

RESULT 36
BG289135
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BG289135 864 bp mRNA linear EST 21-FEB-2001
602387523F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516420 5',
mRNA sequence.
BG289135
BG289135.1 GI:13044673
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 864)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>

Plate: LLAM10407 row: n column: 05
High quality sequence stop: 634.

FEATURES

source

1..864
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4516420"
/issue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bladder; Vector: PCMV-SPORT6; Site 1: NotI; Site 2: SalI; cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."

ORIGIN

Query Match 13.0%; Score 570; DB 4; Length 864;

Best Local Similarity 100.0%; Pred. No. 3.4e-211; Mismatches 0; Indels 0; Gaps 0;

Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1972 GAGAGTGGCAGAGACCTGTATATGTAACCTTGTCTCACAATATGTTATCACTGCTGAT 2031

DB 1 GAGAGTGGCAGAGACCTGTATATGTAACCTTGTCTCACAATATGTTATCACTGCTGAT 60

QY 2032 AATACCTTTCATCTCTCTGCTGCTTTGTTTCATTACTGATTTTCACAAAACCTCTT 2091

DB 61 AATACCTTTCATCTCTCTGCTGCTTTGTTTCATTACTGATTTTCACAAAACCTCTT 120

QY 2092 CATTGGCTAATGTAGATGAGAGGATGATGGGATTCCTTCCCTTTTGGGAAA 2151

DB 121 CATTGGCTAATGTAGATGAGAGGATGATGGGATTCCTTCCCTTTTGGGAAA 180

QY 2152 TGGGCTCTCAAGCTAAAGCTATAGATGAGATTCAGAGTTTCAGGGCTCTGTTCTA 2211

DB 181 TGGGCTCTCAAGCTAAAGCTATAGATGAGATTCAGAGTTTCAGGGCTCTGTTCTA 240

QY 2212 TACATTTGCCCTATGTAAAGGGGTAAAGGGCTCTCTTCAATAGACATGTGAAAGTAA 2271

DB 241 TACATTTGCCCTATGTAAAGGGGTAAAGGGCTCTCTTCAATAGACATGTGAAAGTAA 300

QY 2272 GGAGCCCTTCCCTTGAAGCTGTGCTGATGAGACCTCTCAACCTGGTAAACCTCC 2331

DB 301 GGAGCCCTTCCCTTGAAGCTGTGCTGATGAGACCTCTCAACCTGGTAAACCTCC 360

QY 2332 TTATAGTGGGTATAGATTTTAAACCTTAAATTAACAACAACCTCCATGAGCTT 2391

DB 361 TTATAGTGGGTATAGATTTTAAACCTTAAATTAACAACAACCTCCATGAGCTT 420

QY 2392 TAGAGCCAGAGAGAAATGACAAAGTGAAGGATGAAGCAACCATCTTTCACAGATGAA 2451

DB 421 TAGAGCCAGAGAGAAATGACAAAGTGAAGGATGAAGCAACCATCTTTCACAGATGAA 480

QY 2452 AAGACATCGAGAGCTGTATGATTAACCTGTGAAAAGATGTGTTTGAACATAT 2511

DB 481 AAGACATCGAGAGCTGTATGATTAACCTGTGAAAAGATGTGTTTGAACATAT 540

QY 2512 CTGTGATACAGTATGTGGAGGAGATGTT 2541

DB 541 CTGTGATACAGTATGTGGAGGAGATGTT 570

RESULT 37
BQ187281 723 bp mRNA linear EST 30-Apr-2002
LOCUS BQ187281

DEFINITION UI-E-EJ1-a12-b-15-0-UI-r1 UI-E-EJ1 Homo sapiens cDNA clone

ACCESSION BQ187281

VERSION BQ187281.1 GI:20362832

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 723)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@iowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 REVERSE.
Location/Qualifiers
1..723
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EJ1-a12-b-15-0-UI"
/issue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_1b="UI-E-EJ1"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-EJ1 is a subcloned cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGATTAGAGA; lens, CGATTAGAGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAGTG; retina, CCGGG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

FEATURES

source

ORIGIN

Query Match 12.8%; Score 561; DB 5; Length 723;

Best Local Similarity 100.0%; Pred. No. 1.1e-207; Mismatches 0; Indels 0; Gaps 0;

Matches 561; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 870 GAAAGCGGATGAGCTGTAAAGAGATTAACGAAAGAAAGCAAGCTGTTACG 929

DB 49 GAAAGCGGATGAGCTGTAAAGAGATTAACGAAAGAAAGCAAGCTGTTACG 108

QY 930 GGCTATGAACTGAAGAGAGAGGAAAGAGCTATCTGAGAAATTAACCTGAGTGC 989

DB 109 GGCTATGAACTGAAGAGAGAGGAAAGAGCTATCTGAGAAATTAACCTGAGTGC 168

QY 990 CAGCCGAGCTTCCGAGACATCCGACTCTGCTCCCAAGCCCTTTCATGTGGCCG 1049

DB 169 CAGCCGAGCTTCCGAGACATCCGACTCTGCTCCCAAGCCCTTTCATGTGGCCG 228

QY 1050 AGTGAAGGAGCATTAAGAAATCCCATTTGGAAGTACGAAAGAAAGATCTCGTT 1109

DB 229 AGTGAAGGAGCATTAAGAAATCCCATTTGGAAGTACGAAAGAAAGATCTCGTT 288

QY 1110 AAAAGCTGCTCTCTGAATTTTCAAAAAGTCAAAACAAAACCTCTAAGACTCTCTATT 1169
 DB 289 AAAAGCTGCTCTCTGAATTTTCAAAAAGTCAAAACAAAACCTCTAAGACTCTCTATT 348
 QY 1170 AAAGAGAACCTCTGCT 1229
 DB 349 AAAGAGAACCTCTGCT 408
 QY 1230 GAAACCCCAAGAAAGCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1289
 DB 409 GAAACCCCAAGAAAGCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 468
 QY 1290 AAGGACCCAGACCCATCCCAAGAGAAAGCCCTCTCTCTCTCTCTCTCTCTCTCT 1349
 DB 469 AAGGACCCAGACCCATCCCAAGAGAAAGCCCTCTCTCTCTCTCTCTCTCTCTCT 528
 QY 1350 TACCTTTCCACACAGAAATGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1409
 DB 529 TACCTTTCCACACAGAAATGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 588
 QY 1410 CCATTACGGGAATCTCTCTCA 1430
 DB 589 CCATTACGGGAATCTCTCTCA 609
 RESULT 38
 AM182868/c 558 bp mRNA linear EST 18-NOV-1999
 LOCUS AM182868 xj64h09.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:2662049 3', mRNA sequence.
 ACCESSION AM182868
 VERSION AM182868.1 GI:6451328
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 558)
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 461.
 FEATURES
 source
 1. 558
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2662049"
 /lab_host="DH10B"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBH19W, testis NHT, and B-cell
 NCI CGAP GCHI) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 1 M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Benito
 Soares and M. Fatima Bonaldo. "

Query Match 12.8%; Score 558; DB 2; Length 558;
 Best Local Similarity 100.0%; Pred. No. 1.7e-206;

Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3787 CCATTGCTGTTGGCTTACCCAGAGGAGCTCCAGTTTCTGACTTGAAGTAGACTGAGA 3846
 DB 558 CCATTGCTGTTGGCTTACCCAGAGGAGCTCCAGTTTCTGACTTGAAGTAGACTGAGA 499
 QY 3847 AGAATCCAGAGGCTGCTATCTGCGCAGATTAAAGTAGATTCTATTTCTTGTCTCCCT 3906
 DB 498 AGAATCCAGAGGCTGCTATCTGCGCAGATTAAAGTAGATTCTATTTCTTGTCTCCCT 439
 QY 3907 CTCCTGAGAGACTCTTATTTTATTTGCCCCCTCTCTAGGTTAATCTCTTGTGATTGA 3966
 DB 438 CTCCTGAGAGACTCTTATTTTATTTGCCCCCTCTCTAGGTTAATCTCTTGTGATTGA 379
 QY 3967 CTTTGTGAGAGAGGTTGAGACAGTAGATTAGCAAGTTCCAAAGTCCAAATTAACAGTG 4026
 DB 378 CTTTGTGAGAGAGGTTGAGACAGTAGATTAGCAAGTTCCAAAGTCCAAATTAACAGTG 319
 QY 4027 TGTAGAGTGTGGGGGAAAAATTAGTCTTATTTTCCATGATGGATACACACTGTGA 4086
 DB 318 TGTAGAGTGTGGGGGAAAAATTAGTCTTATTTTCCATGATGGATACACACTGTGA 259
 QY 4087 ATTCAATCTTCAACTGAAAGCCCTGCAAGTCTCTTAAACATAGTGTGTTGTTCTTT 4146
 DB 258 ATTCAATCTTCAACTGAAAGCCCTGCAAGTCTCTTAAACATAGTGTGTTGTTCTTT 199
 QY 4147 AACAAAGTTAAGCTAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4206
 DB 198 AACAAAGTTAAGCTAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 139
 QY 4207 TTGTTTATGCCCCATTTCAATATTTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4266
 DB 138 TTGTTTATGCCCCATTTCAATATTTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 79
 QY 4267 TCTGATGTGTAATAATTTGTTGTCTTGTATTAATCTTAAAGGTTCAATTTGAATAAA 4326
 DB 78 TCTGATGTGTAATAATTTGTTGTCTTGTATTAATCTTAAAGGTTCAATTTGAATAAA 19
 QY 4327 CTATGTGGCTGTAAATA 4344
 DB 18 CTATGTGGCTGTAAATA 1
 RESULT 39
 BI032095/c 621 bp mRNA linear EST 14-JUN-2001
 LOCUS BI032095 PM4-MT0427-230401-003-g07 MT0427 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BI032095
 ACCESSION BI032095
 VERSION BI032095.1 GI:14438725
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 621)
 AUTHORS Dias Neto E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Megal, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F., D.H.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V.,
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Title Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM4&t=PM4-MT0427-230401-003-907&t3=2001-04-23&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 601.

FEATURES

source

1. 621
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="MT0427"
/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 12.7%; Score 556; DB 4; Length 621;

Best Local Similarity 100.0%; Pred.No.1e-205; Indels 0; Gaps 0;

Matches 556; Conservative 0; Mismatches 0;

3535 GGGCTCTCATCTCACAACCTTAAGAGAGATTTCTAGAAAACTGGCCAGATTTCTTT 3594
597 GGGCTCTCATCTCACAACCTTAAGAGAGATTTCTAGAAAACTGGCCAGATTTCTTT 538
3595 GTTCTCATATTATTAATGTGGCAGGCTGTTCAGTTTCTTACTCTTAAGTATAT 3654
537 GTTCTCATATTATTAATGTGGCAGGCTGTTCAGTTTCTTACTCTTAAGTATAT 478
3655 TTCTTGTAACGTCGTAAGAAAAAGAACCAATCAATGTCCTGACTTTGTTCTT 3714
477 TTCTTGTAACGTCGTAAGAAAAAGAACCAATCAATGTCCTGACTTTGTTCTT 418
3715 TGATCCCTCAGTTCTCTTGATTTGACATGTCGGGTTCTTAATTTGGTATAGT 3774
417 TGATCCCTCAGTTCTCTTGATTTGACATGTCGGGTTCTTAATTTGGTATAGT 358
3775 TAGCAATTTAACCATTTGTTGTCCTTACCCAGGGGACTCCCGAGTTTGCAGTTGA 3834
357 TAGCAATTTAACCATTTGTTGTCCTTACCCAGGGGACTCCCGAGTTTGCAGTTGA 298
3835 AGTAGACTGAGAAATCCAGAGGTCATCTGCGGCAATTAAGTAGATTTCTATTC 3894
297 AGTAGACTGAGAAATCCAGAGGTCATCTGCGGCAATTAAGTAGATTTCTATTC 238
3895 TTGGTCTCCCTCTCCCTGAGGACTCTTATTTATTTATTTATTTATTTATTTATTT 3954
237 TTGGTCTCCCTCTCCCTGAGGACTCTTATTTATTTATTTATTTATTTATTTATTT 178
3955 CCTTGAATTTGACTTTGTTGAGAGAGGTTGACAGTAGATTAGCAAAAGTTCCAAAGTGC 4014
177 CCTTGAATTTGACTTTGTTGAGAGAGGTTGACAGTAGATTAGCAAAAGTGC 118
4015 AAAATTTACAGTGTGTAAGATGTGGGGGAAAAATTAATCTTATTTTCCCTACATGGGAT 4074
117 AAAATTTACAGTGTGTAAGATGTGGGGGAAAAATTAATCTTATTTTCCCTACATGGGAT 58
4075 ACAACACTGTGAATTC 4090
57 ACAACACTGTGAATTC 42

RESULT 40
AV722833 629 bp mRNA linear EST 16-OCT-2000
LOCUS AV722833
DEFINITION AV722833 Hrb Homo sapiens cDNA clone HrbAMB11 5', mRNA sequence.

ACCESSION AV722833
VERSION AV722833.1 GI:10825728
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

AUTHORS

Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R., Chen,J., Chen,Z. and Han,Z.

TITLE

Human sapiens cDNA HTB clones

JOURNAL

Unpublished (2000)

COMMENT

Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

source

1. 629
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="HTBAMB11"
/cissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="HTB"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 12.6%; Score 553; DB 1; Length 629;

Best Local Similarity 100.0%; Pred.No.1.5e-204; Indels 0; Gaps 0;

Matches 553; Conservative 0; Mismatches 0;

3463 CTCTAATAGGGAACCTGCTGTAACCTTCAATTCAGCAGAGATGTAGAGAAATAGAGACT 3522
1 CTCTAATAGGGAACCTGCTGTAACCTTCAATTCAGCAGAGATGTAGAGAAATAGAGACT 60
3523 TAATTCACAGAGGCTCTCATCTCAACCTTAAGAGAGATTTCTAGAAAACTGGGC 3582
61 TAATTCACAGAGGCTCTCATCTCAACCTTAAGAGAGATTTCTAGAAAACTGGGC 120
3583 CAGATTTCTTTGTCCTCATCATTTTAATGTCGAGGCTGTTCAATTTCTTACTCTTA 3642
121 CAGATTTCTTTGTCCTCATCATTTTAATGTCGAGGCTGTTCAATTTCTTACTCTTA 180
3643 CCTATGATATTTCTTCTGTAACGTGTCACAAAAAGAAAAAGAACCAATCAGTGTCTTT 3702
181 CCTATGATATTTCTTCTGTAACGTGTCACAAAAAGAAAAAGAACCAATCAGTGTCTTT 240
3703 GACTTTGTTCTTGATCCCTCAGTTTCTTCTGATTTCAAGCATGTGTGGGTTCTTAATT 3762
241 GACTTTGTTCTTGATCCCTCAGTTTCTTCTGATTTCAAGCATGTGTGGGTTCTTAATT 300
3763 TTGGGTAATGAGTAAGCAATTTAACAATGTGTTTGTGCGCTTACCAAGGGAATCCCCAG 3822
301 TTGGGTAATGAGTAAGCAATTTAACAATGTGTTTGTGCGCTTACCAAGGGAATCCCCAG 360
3823 TTCTGACTTGAAGTAGACTGAGAGAAATCCAGAGGTCATCTGAGCCAGATTTAAGTA 3882
361 TTCTGACTTGAAGTAGACTGAGAGAAATCCAGAGGTCATCTGAGCCAGATTTAAGTA 420
3883 GATTCTAATTTCTTGATTTCTCCCTCTCCCTGAGGACTCTTATTTATTTATTTATTTATTT 3942
421 GATTCTAATTTCTTGATTTCTCCCTCTCCCTGAGGACTCTTATTTATTTATTTATTTATTT 480
3943 TAGGTTAATTTCTCTTGAATTTGACTTTGTTGAGAGAGGTTGACAGTAGATTAGCAA 4002

Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: Professor Miklos Palkovits
 cDNA Library Preparation: Michael Brownstein / Ted Uadlin
 Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: NDCM274 row: f column: 09
 High quality sequence stop: 560.

FEATURES

source

Location/Qualifiers
 1..628
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30723296"
 /tissue_type="Human Brain - Frontal Cortex"
 /lab_host="DH10B Tona"
 /clone_lib="NIH_MGC_229"
 /note="Organ: Brain/CNS; Vector: pDNR-LIB; Site 1: Sfil
 (directional); Site 2: Sfil (directional); Library is
 oligo-dr primed and directionally cloned.5' and 3'
 adaptors were used in cloning as follows:
 5'-AACGACGTGATCAGCAGAGTGGCCATTACGCGCGG-3'-ATTCTAGAGG
 CGAGCGCGCCGACATG-d(T)3N-IN-3. Full-length enriched
 library was constructed using the Clontech Creator SMART
 kit and size-selected for >0.5kb with an average insert
 size of 1.2kb library created in the laboratory of
 Jonathan Kuo and Ted Uadlin."

ORIGIN

Query Match 12.6%; Score 551; DB 7; Length 628;
 Best Local Similarity 100.0%; Pred. No. 9.1e-204;
 Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2562 GTTAAAGGATTAACAATTTCTTTACTGTCTCAAAAACCAAACTTTGGAAAAGAAAGTGGG 2621
DB 2 GTTAAAGGATTAACAATTTCTTTACTGTCTCAAAAACCAAACTTTGGAAAAGAAAGTGGG 61
QY 2622 GATGGTTAGTTTACAGAAAGTTACAGCTGTAAACAAAGACCTTAGTATTTGGGATGGC 2681
DB 62 GATGGTTAGTTTACAGAAAGTTACAGCTGTAAACAAAGACCTTAGTATTTGGGATGGC 121
QY 2682 ATGCCAAACCTGTATTAATGTCCTTGTATCACATCACTTCTCAAGTATTCCTTCATTGG 2741
DB 122 ATGCCAAACCTGTATTAATGTCCTTGTATCACATCACTTCTCAAGTATTCCTTCATTGG 181
QY 2742 GCTTCATCCTTTTACAGAACTCTTGTGTGGATGAGACTTACAGAGGATGAGGAGGA 2801
DB 182 GCTTCATCCTTTTACAGAACTCTTGTGTGGATGAGACTTACAGAGGATGAGGAGGA 241
QY 2802 GAGTGGAAATAGTGGCTTCTTTGGCTGGCAAAATGCTACATCTTGGAAAACAAACAGAT 2861
DB 242 GAGTGGAAATAGTGGCTTCTTTGGCTGGCAAAATGCTACATCTTGGAAAACAAACAGAT 301
QY 2862 GTACCTATAGAGCTTCTCAATCACTTGTAAAAATTAATTGTATGTATCAATCTTGGT 2921
DB 302 GTACCTATAGAGCTTCTCAATCACTTGTAAAAATTAATTGTATGTATCAATCTTGGT 361
QY 2922 CCTCTCCCTCCCGTTTGTAAAAATATCAGATAGACATCCAGGCGACCTTGTGTCTCA 2981
DB 362 CCTCTCCCTCCCGTTTGTAAAAATATCAGATAGACATCCAGGCGACCTTGTGTCTCA 421
QY 2982 GTGTAGATCCCTATTAATCTATCGAAGAAATATAGCAGACCTCTGTGTCTCAAT 3041
DB 422 GTGTAGATCCCTATTAATCTATCGAAGAAATATAGCAGACCTCTGTGTCTCAAT 481
QY 3042 ATATAGAAATGCTTCTTATAGCTTACAGACTATTTGTGAACCAATAGAGGATCTA 3101
DB 482 ATATAGAAATGCTTCTTATAGCTTACAGACTATTTGTGAACCAATAGAGGATCTA 541

```

QY 3102 ATCTCCTAGAA 3112
 DB 542 ATCTCCTAGAA 552

RESULT 43
 LOCUS CA447820/c 643 bp mRNA 1linear EST 08-NOV-2002
 DEFINITION UI-H-E10-ayf-d-15-0-UI.s1 NCI CGAP_E10 Homo sapiens cDNA clone
 UI-H-E10-ayf-d-15-0-UI 3', mRNA sequence.

ACCESSION CA447820
 VERSION CA447820.1 GI:24812240
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 643)
 NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov

Tissue Procurement: Dr. Jose Mercuende
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@uiowa.edu
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1..643
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-E10-ayf-d-15-0-UI"
 /tissue_type="Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP E10"
 /note="Organ: Left Pelvis; Vector: p773-Pac (Pharmacia)
 with a modified polylinker; Site 1: EcoR I; Site 2: Not I;
 NCI CGAP_E10 is a cDNA library containing the following
 tissue(s): Chondrosarcoma. The library was constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dr primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into p773-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (drr)18 tail. The
 sequence tag for this library is ACAACTTGAC.
 TAG_TISSUE=chondrosarcoma
 TAG_LIB=UI-H-E10
 TAG_SEQ=ACAACTTGAC"

ORIGIN

Query Match 12.5%; Score 546; DB 6; Length 643;
 Best Local Similarity 100.0%; Pred. No. 8.1e-202;
 Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 3813 GACTCCCGAGTTCTGACTGTAAGTACAGAGAAATCCAGAGTGTATCTGAGCCA 3872
DB 546 GACTCCCGAGTTCTGACTGTAAGTACAGAGAAATCCAGAGTGTATCTGAGCCA 487
QY 3873 GATTTAAGATATTCTATTTCTTGGTCTCCCTCCCTGAGACCTTATTTATTTG 3932
DB 486 GATTTAAGATATTCTATTTCTTGGTCTCCCTCCCTGAGACCTTATTTATTTG 427
QY 3933 TCCCTCTCTTAGGTTAATTCCTTTGATTTGATTGTAAGAAAGAGTTGACACT 3992

```


|||||
Db | 426 | TCCCTCTTTCAGGTTAATCTCCTTGTGATTGACTTGTGTGAGAGAGAGTTGAGACGT | 367
Qy | 3993 | AGATTAGCAAGTCCCAAGTGCACAAATTAACAGTGTGTAGAGTGTGGGGGAAAATTAGT | 4052
Db | 366 | AGATTAGCAAGTCCCAAGTGCACAAATTAACAGTGTGTAGAGTGTGGGGGAAAATTAGT | 307
Qy | 4053 | CTATATTTTCCCTACATGAGATACACAGCTGTGAATTCATCTTCAACTGAGAGCCCTGC | 4112
Db | 306 | CTATATTTTCCCTACATGAGATACACAGCTGTGAATTCATCTTCAACTGAGAGCCCTGC | 247
Qy | 4113 | AGTTCTCTTAAACATATGTTGTTTCTTTTAAACAAGTTTAAAGTACTGTTAATTA | 4172
Db | 246 | AGTTCTCTTAAACATATGTTGTTTCTTTTAAACAAGTTTAAAGTACTGTTAATTA | 187
Qy | 4173 | ATTAATAAATAATGCTGTGCTACTGCTTGTGTTTAAAGCCATTCATATGTT | 4232
Db | 186 | ATTAATAAATAATGCTGTGCTACTGCTTGTGTTTAAAGCCATTCATATGTT | 127
Qy | 4233 | GTCGTGTGTATTCATTAATCTTTTGTATACCATTTGTATGTAATAATTGTTGCTTG | 4292
Db | 126 | GTCGTGTGTATTCATTAATCTTTTGTATACCATTTGTATGTAATAATTGTTGCTTG | 67
Qy | 4293 | TAAATATCTTAATAAGATTCATTTGTAATTAATCTATTTGCTGTTAAAAA | 4352
Db | 66 | TAAATATCTTAATAAGATTCATTTGTAATTAATCTATTTGCTGTTAAAAA | 7
Qy | 4353 | AAAAAA 4358
Db | 6 | AAAAAA 1

RESULT 44
BM989342/c 544 bp mRNA linear EST 17-JUN-2002
LOCUS BM989342
DEFINITION UI-H-DP0-at8-d-08-0-UI.s1 NCI CGAP_Fs1 Homo sapiens cDNA clone
IMAGE:5863639 3', mRNA sequence.
BM989342
VERSION BM989342.1 GI:19708731
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 544)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Mary Hendrix
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Seq Primer: M13 FORWARD
POLYA=yes.

FEATURES
Source
1. 544
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5863639"
/tissue_type="Fibrosarcoma"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Fsi"
/note="Vector: pT73-Pac (pharmacia) with a modified
polylinker. Site_1: EcoR I. Site_2: Not I; NCI CGAP_Fs1 is
a cDNA library containing the following tissue(s):
Fibrosarcoma Cell line HT-108 (ATCC number CCL-121). The
library was constructed according to Bonaldo, Lennon and

Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GTTCTACGAG.
TAG TISSUE=fibrosarcoma
TAG LIB=UI-H-DP0
TAG_SEQ=GTTCTACGAG"

ORIGIN
Query Match 12.4%; Score 544; DB 5; Length 544;
Best Local Similarity 100.0%; Pred. No. 5e-201;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy | 3815 | CTCCCAAGTTTCTGACTTGAAGTAGACTGAGAAAGATCCAGAGTCTATCTGGCCAGA | 3874
Db | 544 | CTCCCAAGTTTCTGACTTGAAGTAGACTGAGAAAGATCCAGAGTCTATCTGGCCAGA | 485
Qy | 3875 | TTTAAGTAGATTCATTTCTGTTGTTCTCCCTCTCCCTGAGACCTTTATTTATGTC | 3934
Db | 484 | TTTAAGTAGATTCATTTCTGTTGTTCTCCCTCTCCCTGAGACCTTTATTTATGTC | 425
Qy | 3935 | CCTCTCTAGGTTAATCTCTTGAATTTGACTTGTGAGAGAGAGTTGACAGTAG | 3994
Db | 424 | CCTCTCTAGGTTAATCTCTTGAATTTGACTTGTGAGAGAGAGTTGACAGTAG | 365
Qy | 3995 | ATTAGCAAGTTCACAGTGCACAAATTAACAGTGTGTAGAGTGTGGGGGAAAATTAGTCT | 4054
Db | 364 | ATTAGCAAGTTCACAGTGCACAAATTAACAGTGTGTAGAGTGTGGGGGAAAATTAGTCT | 305
Qy | 4055 | TATTTTCCCTACATGAGATCAACACTGTGAATTCATCTTCACTGAGAGCCCTGACG | 4114
Db | 304 | TATTTTCCCTACATGAGATCAACACTGTGAATTCATCTTCACTGAGAGCCCTGACG | 245
Qy | 4115 | TTCTCCCAAAACATGATGTTGTTTCTTTTAAACAAGTTTAAAGTGAATTAAT | 4174
Db | 244 | TTCTCCCAAAACATGATGTTGTTTCTTTTAAACAAGTTTAAAGTGAATTAAT | 185
Qy | 4175 | TAAAAAATATGCTGTCTGTACTACTTCACTTGTGTTTAAAGCCATTCATATGTTGT | 4234
Db | 184 | TAAAAAATATGCTGTCTGTCTGTACTACTTCACTTGTGTTTAAAGCCATTCATATGTTGT | 125
Qy | 4235 | CTGTGTGTATTCATTAATCTTTGATACATTTCTGTGATGTAAATGTTGTTGTA | 4294
Db | 124 | CTGTGTGTATTCATTAATCTTTGATACATTTCTGTGATGTAAATGTTGTTGTA | 65
Qy | 4295 | AATATCTTAATAAGATTCATTTGTAATAATCTATTTGTGGCTTTAAAAA | 4354
Db | 64 | AATATCTTAATAAGATTCATTTGTAATAATCTATTTGTGGCTTTAAAAA | 5
Qy | 4355 | AAAAA 4358
Db | 4 | AAAAA 1

RESULT 45
BM729196 567 bp mRNA linear EST 01-MAR-2002
LOCUS BM729196
DEFINITION UI-E-B01-aiv-n-03-0-UI.r1 UI-E-B01 Homo sapiens cDNA clone
IMAGE:5863639 3', mRNA sequence.
BM729196
VERSION BM729196.1 GI:19050529
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 567)
Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 8889548

COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com).
Seq primer: M13 Reverse.

FEATURES

source

1..567

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-E-EOL-aiw-n-03-0-UI"

/issue_type="fetal eye"

/dev_stage="fetal"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-EOL is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGGTATAC. This library was created for the program Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Query Match 12.4%; Score 542; DB 4; Length 567;
Best Local Similarity 100.0%; Pred. No. 3e-200;
Matches 542; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2749 CCTTTAGCAGAACTCTTGTGTGGATAGAGACTTGGAGGGGAGAGAGTGTG 2808
DB 1 CTTTATAGCAGAACTCTTGTGTGGATAGAGACTTGGAGGGGAGAGAGTGTG 60

QY 2809 GAAATAGTGCTCTCTTGGCTGGCAATGTCTACATCTTGAACAAACAGATGTACTTA 2868
DB 61 GAAATAGTGCTCTCTTGGCTGGCAATGTCTACATCTTGAACAAACAGATGTACTTA 120

QY 2869 ATGAGGCTCCATCCTTGTAAATAATTTGTAATGTAACATCTTGGTCTCTCC 2928
DB 121 ATGAGGCTCCATCCTTGTAAATAATTTGTAATGTAACATCTTGGTCTCTCC 180

QY 2929 CTTCCGCTTTTGAATAATACAGATAGACCTCCAGGCACTTGGTCTCAGTGTAG 2988
DB 181 CTTCCGCTTTTGAATAATACAGATAGACCTCCAGGCACTTGGTCTCAGTGTAG 240

QY 2989 ATCCCTATTACTATCGAAAGAAATAGACCAAGACCTTGGTCTCAATATATAGG 3048
DB 241 ATCCCTATTACTATCGAAAGAAATAGACCAAGACCTTGGTCTCAATATATAGG 300

QY 3049 AATTGCTTTCTTAACTTCAGACATATGTGTGAAGAAAGAGAGGGGTCTATCTCCT 3108
|||||

DB 301 AATTGCTTTCTTAACTTCAGACATATGTGTGAAGAAAGAGAGGGGTCTATCTCCT 3160

QY 3109 AGAAGGTAGGGGCTTTATCTTTAAAGAGAAATATGTCCTCCCATTTATGCACTTTTAA 3168
|||||

DB 361 AGAAGGTAGGGGCTTTATCTTTAAAGAGAAATATGTCCTCCCATTTATGCACTTTTAA 420
|||||

QY 3169 GGAGAGCCAGGATATGATGAGGTGTGTGCTGGCCCATCATGATGACACAGAGAGAAAT 3228
|||||

DB 421 GGAGAGCCAGGATATGATGAGGTGTGTGCTGGCCCATCATGATGACACAGAGAGAAAT 480
|||||

QY 3229 GGATATCCATTGTGGAGAGAGAAAGATTCTTCAGGGGCTTCCCATGCTTAAAGTTT 3288
|||||

DB 481 GGATATCCATTGTGGAGAGAGAAAGATTCTTCAGGGGCTTCCCATGCTTAAAGTTT 540
|||||

QY 3289 TT 3290

DB 541 TT 542

RESULT 46

BM852930 540 bp mRNA linear EST 06-MAR-2002
LOCUS K-EST0134254 S22SNUI6n1 Homo sapiens cDNA clone S22SNUI6n1-130-C06
DEFINITION 5', mRNA sequence.
ACCESSION BM852930
VERSION BM852930.1 GI:19209329
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 540)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
CONTACT: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boseun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongseung@mail.kribb.re.kr
Plate: 130 row: C column: 06
High quality sequence stop: 540.
Location/Qualifiers

FEATURES

source

1..540

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="S22SNUI6n1-130-C06"

/sex="F"

/issue_type="Ascites"

/cell_line="SNUI-16"

/lab_host="DH10B"

/lab_host="SNUI-16"

/note="Organ: Stomach; Vector: p773-Pac; Site 1: EcoRI; Site 2: NotI. The S22SNUI6 library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cells of SNUI-16 culture. SNUI-16 cell was obtained from Korean Cell Line Bank (KCLB). SNUI-16 was established from ascitic fluids of Korean patients by Park J.G. et al. (1990), Cancer Res 50: 2773-2780."

ORIGIN

Query Match 12.4%; Score 540; DB 4; Length 540;
Best Local Similarity 100.0%; Pred. No. 1.8e-199;
Matches 540; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2072 TGATTCACAAAACTCTTTCATTGCGCTAATGTGATGATTAGAGGGTGAATTT 2131
DB 1 TGATTCACAAAACTCTTTCATTGCGCTAATGTGATGATTAGAGGGTGAATTT 60
QY 2132 CTTTCCCTTTTGGAAATGGGCTCTCAAGCTTAAGCTTAAGGATGCGAATTCGAA 2191
DB 61 CTTTCCCTTTTGGAAATGGGCTCTCAAGCTTAAGCTTAAGGATGCGAATTCGAA 120
QY 2192 GTTTCAGGGGCTGTCTTCTTATACATTTGCTATGTTAAAGGGGCTCTCTCA 2251
DB 121 GTTTCAGGGGCTGTCTTCTTATACATTTGCTATGTTAAAGGGGCTCTCTCA 180
QY 2252 TTAGACATGTGAAGATGAAGACAGCCCTTCTTTAGAGCTGTGCTCATGGCACTTT 2311
DB 181 TTAGACATGTGAAGATGAAGACAGCCCTTCTTTAGAGCTGTGCTCATGGCACTTT 240
QY 2312 CTACACCTGTACACCTCTCTTATATGTGGGTATATGATTTTAACTTAAATTAACA 2371
DB 241 CTACACCTGTACACCTCTCTTATATGTGGGTATATGATTTTAACTTAAATTAACA 300
QY 2372 AACACCTCACATGAGCTTTAGAGCCAGAAAGAAATGACAAGTAAAGGATGAAGCAA 2431
DB 301 AACACCTCACATGAGCTTTAGAGCCAGAAAGAAATGACAAGTAAAGGATGAAGCAA 360
QY 2432 GCCATCTTCACAGATGAAGAAAGCATCGAGAGTTGATGATTAATCTGTGAAGATA 2491
DB 361 GCCATCTTCACAGATGAAGAAAGCATCGAGAGTTGATGATTAATCTGTGAAGATA 420
QY 2492 GTTGTTCATTGAACTATCTGTGATACATGTCATGTGGAGAGGATGTTGGCTGTGAT 2551
DB 421 GTTGTTCATTGAACTATCTGTGATACATGTCATGTGGAGAGGATGTTGGCTGTGAT 480
QY 2552 TATTTTTCAGTAAATGATTAACAATTTCTTACTGTCAAAAACCAAAATCTTTGAAA 2611
DB 481 TATTTTTCAGTAAATGATTAACAATTTCTTACTGTCAAAAACCAAAATCTTTGAAA 540

RESULT 47
AA908959 536 bp mRNA linear EST 09-JUN-1998
LOCUS AA908959
DEFINITION IMAGE:1523095 3', mRNA sequence.
ACCESSION AA908959
VERSION AA908959.1 GI:3048364
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euteleostomi; Primates; Catarrhini; Homnidae; Homo.
TITLE NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Unpublished (1997)
Email: cgabs-remail.nih.gov
Contact: Robert Strausberg, Ph.D.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 533 Std Error: 0.00
Seq primer: -40ml3 fwd. RT from Amersham
High quality sequence stop: 482.
Location/Qualifiers
1. 536
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1523095"
/lab_host="DH10B"
/clone_lib="soares_NFL_T_GBC_S1"
/note="Organ: pooled; Vector: pTTT3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH19W, testis NHT, and B-cell

NCI-CCGAP (GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.

ORIGIN

Query Match 12.3%; Score 536; DB 1; Length 536;
Best Local Similarity 100.0%; Pred. No. 6,6e-198;
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3813 GACTCCCAAGTTCTGACCTGAAGTACAGTGAAGATCCAGAGGTCTATCTGCGCA 3872
DB 536 GACTCCCAAGTTCTGACCTGAAGTACAGTGAAGATCCAGAGGTCTATCTGCGCA 477
QY 3873 GATTTAGATGATTCATTTCTTGGTTCCTCTCCCTGAGACCTTATTTTATTTG 3932
DB 476 GATTTAGATGATTCATTTCTTGGTTCCTCTCCCTGAGACCTTATTTTATTTG 417
QY 3933 TCCCTCTTCTAGATTAATCTCTTTGATTTGATCTTTGTAAGAGAGTTGACAGT 3992
DB 416 TCCCTCTTCTAGATTAATCTCTTTGATTTGATCTTTGTAAGAGAGTTGACAGT 357
QY 3993 AGATTAGCAAGTTCCAGAGCAAAATTAACAGTGTGTAGAGTGGGGGAAATTAAGT 4052
DB 356 AGATTAGCAAGTTCCAGAGCAAAATTAACAGTGTGTAGAGTGGGGGAAATTAAGT 297
QY 4053 CTATATTTTCCCTACATGAGGATCAACACACTGTGAATTCATCTTCACTGAAGCCCTGC 4112
DB 296 CTATATTTTCCCTACATGAGGATCAACACACTGTGAATTCATCTTCACTGAAGCCCTGC 237
QY 4113 AGTTCTCTTAAACATAGTGTGTTGTTTCTTTAAACAAAGTTTAAAGTGTATTA 4172
DB 236 AGTTCTCTTAAACATAGTGTGTTGTTTCTTTAAACAAAGTTTAAAGTGTATTA 177
QY 4173 ATTAATAAAATGCTGTGCTGTCTACCTGCTTGTATGATTCATTAATGTT 4232
DB 176 ATTAATAAAATGCTGTGCTGTCTACCTGCTTGTATGATTCATTAATGTT 117
QY 4233 GTCGTGTTGTAATTCATTAATTCATTTCTGATGTGAATTTGTTGCTTG 4292
DB 116 GTCGTGTTGTAATTCATTAATTCATTTCTGATGTGAATTTGTTGCTTG 57
QY 4293 TAAATATCTTATAAAGATCAATTTGTAATTAATCTATGTGCTTTAAAAA 4348
DB 56 TAAATATCTTATAAAGATCAATTTGTAATTAATCTATGTGCTTTAAAAA 1

RESULT 48
BG547204 536 bp mRNA linear EST 04-APR-2001
LOCUS BG547204
DEFINITION 602574616F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4703066 5',
mRNA sequence.
ACCESSION BG547204
VERSION BG547204.1 GI:13545869
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euteleostomi; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.

Qy	2304	GCACCTCTTCACACCCCTGTATACACCCCTCCTTATATATGGTATATGATATTTTAAACCCAAA	2363
Db	301	GCACCTCTTCACACCCCTGTATACACCCCTCCTTATATATGGTATATGATATTTTAAACCCCTAAA	360
Qy	2364	ATAAAGCAAAACAACCTCAACCATGAGCTTTAGAGACACAGAAAGAGATGCAAGTGAAGCGA	2423
Db	361	ATAAAGCAAAACAACCTCAACCATGAGCTTTAGAGACACAGAAAGAGATGCAAGTGAAGCGA	420
Qy	2424	TGAAGCAAGCCATCTTCAACAGATGTAAGAAAAGACATCGGAGATTGGTAGATTAAGTCTCTG	2483
Db	421	TGAAGCAAGCCATCTTCAACAGATGTAAGAAAAGACATCGGAGATTGGTAGATTAAGTCTCTG	480
Qy	2484	AAAAGATAGTGTTCATTTGAAACTATTTCTGTGTATACATCATATGGAAGAGATG	2539
Db	481	AAAAGATAGTGTTCATTTGAAACTATTTCTGTGTATACATCATATGGAAGAGATG	536
RESULT 50			
BI457952			
LOCUS	BI457952	623 bp	mRNA linear EST 21-AUG-2001
DEFINITION	603198945F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5278546 5', mRNA sequence.		
ACCESSION	BI457952		
VERSION	BI457952.1 GI:15248608		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs-remail.nih.gov Tissue Procurement: Miklos Pavlovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: LLM11703 row: 1 column: 11 High quality sequence stop: 623.		
FEATURES	location/Qualifiers		
SOURCE	1..623		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5278546"		
	/tissue_type="hypothalamus"		
	/lab_host="DH10B"		
	/clone_lib="NIH_MGC_96"		
	/note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3', size-selected for average insert size 2.3 kb and normalized to ROF 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."		
ORIGIN			
Query Match	12.3%	Score 536;	DB 4; Length 623;
Best Local Similarity	99.8%;	Pred. No. 6.4e-198;	
Matches 586;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
Qy	867	ATGGAAGGGCGGATGACGCTGTAAAGAAAGATTAACGAGAAAGAAAGCACAAGCTGTT	926
Db	6	ATGGAAGGGCGGATGACGCTGTAAAGAAAGATTAACGAGAAAGAAAGCACAAGCTGTT	65

QY	92	CAGGGCTATGAACTGTGAAGAGACAGAGAAACAGACTGATCTGTGAAATTTAACTGTGAG	986
Db	66	CAGGGCTATGAACTGTGAAGAGAGAGAAACAGAGCTTACTGTGAAATTTAACTGTGAG	125
QY	987	TGCGAGCCGAGAGCTTTCCGAGACATCCAGACTGTGCTCCAGCCCTTCTCATGTGGG	1046
QY	126	TGCGAGCGGAGAGCTTTCCGAGACATCCAGACTGTGCTCCAGAGCCCTTCTCATGTGGG	185
Db	1047	CGAGATGTGAAAGGACATTAAGAGAAATCCCATTTTGGAAAGTACAGAAAGAAAGACTTCT	1106
Db	186	CGAGATGTGAAAGGACATTAAGAGAAATCCCATTTTGGAAAGTACAGAAAGAAAGACTTCT	245
QY	1107	GTTAAAGAGCTGGCTGCTGTAATTTTCAAAGTCAAAACAAAACTCCTAAGCACTTCTCT	1166
Db	246	GTTAAAGAGCTGGCTGCTGTAATTTTCAAAGTCAAAACAAAACTCCTAAGCACTTCTCT	305
QY	1167	ATTAAAGAGAACCTGTGTGGTTCCTTATCTGAAACGTGTTGTAAACGTGAATTGAGAGC	1286
Db	306	ATTAAAGAGAACCTGTGTGGTTCCTTATCTGAAACGTGTTGTAAACGTGAATTGAGAGC	365
QY	1227	CAAGAAACCCCAAGAAAGCCCGGTCTTTCAGTGGACACCCCAACAATCTCCACTCC	1286
Db	366	CAAGAAACCCCAAGAAAGCCCGGTCTTTCAGTGGACACCCCAACAATCTCCACTCC	425
QY	1287	CAAAAGGAGCCCAAGACCCATCCCAAGAGAAAGCTTCTCAAGTGAATGAGATTTTG	1346
Db	426	CAAAAGGAGCCCAAGACCCATCCCAAGAGAAAGCTTCTCAAGTGAATGAGATTTTG	485
QY	1347	CCGTACCTTTCCACACAGAAATGTATTTGTGTGTGGACACGCTCCCCATCACCG	1406
Db	486	CCGTACCTTTCCACACAGAAATGTATTTGTGTGTGGACACGCTCCCCATCACCG	545
QY	1407	TTTACATTACGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG	1453
Db	546	TTTACATTACGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG	592
RESULT 51			
B0636732			
LOCUS			
DEFINITION	B0636732	608 bp	mRNA
DESCRIPTION	hnl13g06.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he		
ACCESSION	B0636732		
VERSION	B0636732		
KEYWORDS	B0636732.1	GI:21761191	
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 608)		
REFERENCE	Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A.,		
REFERENCE	Touchman,J.W., Boultard,G., Smith,D. and Peterson,K.		
REFERENCE	Expressed sequence tag analysis of human retina for the NEIBank		
REFERENCE	Project: Rebinding, an abundant, novel retinal cDNA and alternative		
REFERENCE	splicing of other retina-preferred gene transcripts		
REFERENCE	Mol. Vis. 8 (4), 196-204 (2002)		
JOURNAL	22103461		
MEDLINE	12107411		
COMMENT	Contact: Wistow G		
COMMENT	Section on Molecular Structure and Function		
COMMENT	National Eye Institute		
COMMENT	6/331, NIH, Bethesda, MD 20892-2740, USA		
COMMENT	Tel: 301 402 3452		
COMMENT	Fax: 301 496 0078		
COMMENT	Email: graeme@nei.nih.gov		
COMMENT	Plate: 13 row: 9 column: 06		
COMMENT	Seq primer: M13RPI reverse primer (AB1).		
COMMENT	Location/Qualifiers		
COMMENT	1..608		
FEATURES			
source	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		

QY 990 CAGCGGAGCTTTCCGAGACATCCAGACTTCGCTCCAGGCCCTTCTCATGTGGCGG 1049
DB 595 CAGCGGAGCTTTCCGAGACATCCAGACTTCGCTCCAGGCCCTTCTCATGTGGCGG 536
QY 1050 AGTGGAAAGGAGCACTAAAGAAATCCCATTTTGAAGTACAGAAAGAAAGACTCTGTT 1109
DB 535 AGTGGAAAGGAGCACTAAAGAAATCCCATTTTGAAGTACAGAAAGAAAGACTCTGTT 476
QY 1110 AAAAGCTGGCTCTGTAATTTTCAAAAGTCAAAACAAAACTCTTAAGCACTCTCTATT 1169
DB 475 AAAAGCTGGCTCTGTAATTTTCAAAAGTCAAAACAAAACTCTTAAGCACTCTCTATT 416
QY 1170 AAAGAGAACCTGTGCTCTTATCTGAAACGTGTTTAAACGTGAATTGAGAGCAA 1229
DB 415 AAAGAGAACCTGTGCTCTTATCTGAAACGTGTTTAAACGTGAATTGAGAGCAA 356
QY 1230 GAAACCCCAAGAAAGCCCGGCTCTTCAGTGACACCCCAAGCACTCTCACTCCCAA 1289
DB 355 GAAACCCCAAGAAAGCCCGGCTCTTCAGTGACACCCCAAGCACTCTCACTCCCAA 296
QY 1290 AAGGACCCCAAGCACTCTTCAAGAGAAAGCTTCAAGTGAATGAAATTTGCGG 1349
DB 295 AAGGACCCCAAGCACTCTTCAAGAGAAAGCTTCAAGTGAATGAAATTTGCGG 236
QY 1350 TACCTTCCACACAGAAATGATTTGTGCTTGACACAGCTCCCATCACCCTTA 1409
DB 235 TACCTTCCACACAGAAATGATTTGTGCTTGACACAGCTCCCATCACCCTTA 176
QY 1410 CCATTACGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG 1453
DB 175 CCATTACGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG 132

RESULT 53
LOCUS AW965141 686 bp mRNA linear EST 01-JUN-2000
DEFINITION EST377214 MAGB resequences, MAGI Homo sapiens cDNA, mRNA sequence.
ACCESSION AW965141
VERSION AW965141.1 GI:8154977
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Hegde, P., Qi, R., Adermally, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 210
Seq primer: Reverse.
FEATURES
source location/Qualifiers
1..686
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_id="MAGE resequences, MAGI"
/note="Vector: pBluescriptSKm"

ORIGIN
Query Match 12.2%; Score 532; DB 2; Length 686;
Best Local Similarity 99.8%; Pred. No. 2,3e-196;
Matches 582; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 760 TGCATTTGACCTTCATGAAACAGACAGACAGACTGCGAGCCCAAGAAAGAGATCG 819
DB 1 TGCATTTGACCTTCATGAAACAGACAGACAGACTGCGAGCCCAAGAAAGAGATCG 60
QY 820 AGGAGCTGAAGTACAGAGAGACACGCTCTTGTGGATTGAAAGTATGAAAGAGGGA 879
DB 61 AGGAGCTGAAGTACAGAGAGACACGCTCTTGTGGATTGAAAGTATGAAAGAGGGA 120
QY 880 TGCAGCTGTAAAGAAAGATACAGAAAGAAAGGCAAGCTGTTTCAAGGCTATGAAA 939
DB 121 TGCAGCTGTAAAGAAAGATACAGAAAGAAAGGCAAGCTGTTTCAAGGCTATGAAA 180
QY 940 CTGAAGAGAGAGAGAAACAGAGCTATCTGAGAAATTTAACTGAGTCCAGCGGAGC 999
DB 181 CTGAAGAGAGAGAGAAACAGAGCTATCTGAGAAATTTAACTGAGTCCAGCGGAGC 240
QY 1000 TTTCCGAGACATCCCAAGCTCTGCTCCCAAGCCCTTCTCATGTGGCGGAGTGGAAAG 1059
DB 241 TTTCCGAGACATCCCAAGCTCTGCTCCCAAGCCCTTCTCATGTGGCGGAGTGGAAAG 300
QY 1060 GACATTAAGAGAAATCCCAATTTGGAAGTACAGAAAGAAAGACTCTGTTAAAGCTGG 1119
DB 301 GACATTAAGAGAAATCCCAATTTGGAAGTACAGAAAGAAAGACTCTGTTAAAGCTGG 360
QY 1120 CTCTGAATTTTCAAAAGTCAAAACAAAACTCTTAAGCACTCTCTATTAAAGAGAAC 1179
DB 361 CTCTGAATTTTCAAAAGTCAAAACAAAACTCTTAAGCACTCTCTATTAAAGAGAAC 420
QY 1180 CTGTGCTTCTTATCTGAAACGTGTTTGTAAAGTGAATGAGAGCCAAAGAAAGCCGAG 1239
DB 421 CTGTGCTTCTTATCTGAAACGTGTTTGTAAAGTGAATGAGAGCCAAAGAAAGCCGAG 480
QY 1240 AAAAGCCCGGCTCTTCAAGTGAACCCCAAGCAAGCTCTCACTCCCAAGAGAGCCCA 1299
DB 481 AAAAGCCCGGCTCTTCAAGTGAACCCCAAGCAAGCTCTCACTCCCAAGAGAGCCCA 540
QY 1300 GCAACCATCCCAAGAGAAAGCTTCTCAAGTGAATGAGAGAGAGAGAGAGAGAGAG 1342
DB 541 GCAACCATCCCAAGAGAAAGCTTCTCAAGTGAATGAGAGAGAGAGAGAGAGAGAG 583

RESULT 54
LOCUS BE870003 531 bp mRNA linear EST 20-OCT-2000
DEFINITION 601449751.F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3853452 5', mRNA sequence.
ACCESSION BE870003
VERSION BE870003.1 GI:10318779
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://img.ncbi.nlm.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.jnl.gov
Plate: LHAM9577 row: f column: 13
High quality sequence stop: 531.
FEATURES
source location/Qualifiers
1..531
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

/clone="IMAGE:3853452"
/issue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NH10 MG6 65"
/note="Organ: colon; Vector: PCMV-SPOrt6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

ORIGIN

Query Match 12.1%; Score 531; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 5.8e-196;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2971 CTTTGGTCTCAGTGTAAAGATCCCTATTACTATCTGAAAGAAATAGACCAAGACCTC 3030
DB 1 CTTTGGTCTCAGTGTAAAGATCCCTATTACTATCTGAAAGAAATAGACCAAGACCTC 60
QY 3031 TGGTCTCAATATATATAGGAATTCGCTTTTGTAGTCTTCAGACTATGTGTGAAACAA 3090
DB 61 TGGTCTCAATATATATAGGAATTCGCTTTTGTAGTCTTCAGACTATGTGTGAAACAA 120
QY 3091 GTAGGGGTCTAATCTCTAGAAAGTAGGGGCTTTTATCTTAAAGAAATATGTCCCAAG 3150
DB 121 GTAGGGGTCTAATCTCTAGAAAGTAGGGGCTTTTATCTTAAAGAAATATGTCCCAAG 180
QY 3151 ATTATTAGCACTTTTAGAGAGAAAGCAAGATATGTAGGGTGTGGCTGGCCATCAGT 3210
DB 181 ATTATTAGCACTTTTAGAGAGAAAGCAAGATATGTAGGGTGTGGCTGGCCATCAGT 240
QY 3211 GGAGCAGAAAGAGATGGGATACATGTGGGAAAGAAAGAAAGTCTCAGAGGGCC 3270
DB 241 GGAGCAGAAAGAGATGGGATACATGTGGGAAAGAAAGAAAGTCTCAGAGGGCC 300
QY 3271 TCCCACTGCTAAAGTTTTTGTAGATGTGATCTGTGCTTCCGTGATTTGACTTTTAA 3330
DB 301 TCCCACTGCTAAAGTTTTTGTAGATGTGATCTGTGCTTCCGTGATTTGACTTTTAA 360
QY 3331 GGAATTAATCTGGCAGACATGTAGTATTTCTTGATGATCTTCTGCTTATTTCTCT 3390
DB 361 GGAATTAATCTGGCAGACATGTAGTATTTCTTGATGATCTTCTGCTTATTTCTCT 420
QY 3391 TTGT 3450
DB 421 TTGT 480
QY 3451 GGAGAGTGGGCTCTCTATAAGGAGACCTGCTGTAACCTTCAATGACAGCAAG 3501
DB 481 GGAGAGTGGGCTCTCTATAAGGAGACCTGCTGTAACCTTCAATGACAGCAAG 531

RESULT 55
CA133331/c 753 bp mRNA linear EST 04-NOV-2002
LOCUS UI-CF-FNO-ay-g-20-0-UI.81 UI-CF-FNO Homo sapiens cDNA clone
DERIVATION UI-CF-FNO-ay-g-20-0-UI 3', mRNA sequence.
ACCESSION CA133331
VERSION CA133331.1 GI:24531429
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 753)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548

COMMENT Contact: McCray, PB
McCrack Lab
University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 1-28, >AT rich#low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES

source
Location/Qualifiers
1..753
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/issue_type="Human Lung Epithelial cells"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-FNO"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-FNO is a subtracted cDNA library derived from two
normalized Human lung epithelial cell libraries (EN1 and
DU1) The library was subtracted according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. For additional information, contact:
bento-soares@uiowa.edu
TAG TISSUE=Human Lung Epithelial Cell lines untreated LPS
6hr to LPS 24h
TAG LIB=UI-CF-FNO
TAG_SEQ=CTGCTCAGGT"

ORIGIN

Query Match 12.1%; Score 530; DB 6; Length 753;
Best Local Similarity 100.0%; Pred. No. 1.4e-195;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3813 GACTCCCACTTCTGACCTTGAAGTACGAGCAAGATCCAGAGTGTATCTGCGCA 3872
DB 554 GACTCCCACTTCTGACCTTGAAGTACGAGCAAGATCCAGAGTGTATCTGCGCA 495
QY 3873 GATTTAAGTAGATTTATTTCTTCTGCTTCCCTCCCTGAGAGACCTCTTATTTATTG 3932
DB 494 GATTTAAGTAGATTTATTTCTTCTGCTTCCCTCCCTGAGAGACCTCTTATTTATTG 435
QY 3933 TCCCTCTTCTAGGTAATTTCTCTTGAATTTGACTTTGTGAGAAAGAGTTGACAGT 3992
DB 434 TCCCTCTTCTAGGTAATTTCTCTTGAATTTGACTTTGTGAGAAAGAGTTGACAGT 375
QY 3993 AGATTAGCAAGTTCCAGAGTCAAAATTTACAGTGTGTAGAGTGTGGGGGAAATTAAGT 4052
DB 374 AGATTAGCAAGTTCCAGAGTCAAAATTTACAGTGTGTAGAGTGTGGGGGAAATTAAGT 315
QY 4053 CTTATTTTCCCTAGACAGTGGATACACACGTGTAATTCATCTTCAATGAGAGCCCTGC 4112
DB 314 CTTATTTTCCCTAGACAGTGGATACACACGTGTAATTCATCTTCAATGAGAGCCCTGC 255
QY 4113 AGTTCTCTAAACATAGTGTGTTGTTTCTTTTAAACAAAGTTTAAAGCTAGTTATTA 4172
DB 254 AGTTCTCTAAACATAGTGTGTTGTTTCTTTTAAACAAAGTTTAAAGCTAGTTATTA 195
QY 4173 ATTAAAAAAATTTGCTGTCTGTCTACTTCACTTCACTTTTATGCCATTTCATATTGT 4232
DB 194 ATTAAAAAAATTTGCTGTCTGTCTACTTCACTTCACTTTTATGCCATTTCATATTGT 135
QY 4233 GTCTGTGTGTATTCATATCACTTTGATACCAATTTCTGANGTGTAAATTTGGTGTCTTG 4292
DB 134 GTCTGTGTGTATTCATATCACTTTGATACCAATTTCTGANGTGTAAATTTGGTGTCTTG 75

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	ORIGIN
4293	4444700/c	696 bp mRNA linear EST 08-NOV-2002	CA444700											
Db	74	TTAATATCTTATAAGAGCTTCAATTGTAAATTAACCTATGTGGCTGTAA	696 bp mRNA linear EST 08-NOV-2002	CA444700										
		UI-H-DHI-aww-h-01-0-U s1 NCI CGAP_DHI Homo sapiens cDNA clone	CA444700											
		UI-H-DHI-aww-h-01-0-U 3', mRNA sequence.	CA444700.1	GI:24809120										
		EST.												
		Homo sapiens (human)												
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;												
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.												
		1 (bases 1 to 696)												
		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .												
		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),												
		Tumor Gene Index												
		Unpublished (1997)												
		Contact: Robert Strausberg, Ph.D.												
		Email: cgapbs-remail.nih.gov												
		Tissue Procurement: Dr. Jose Mercuende												
		cDNA library preparation: Dr. M. Bento Soares, University of Iowa												
		cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa												
		DNA Sequencing by: Dr. M. Bento Soares, University of Iowa												
		Clone Distribution: Clone distribution information can be obtained												
		from Dr. M. Bento Soares, bento-soares@uiowa.edu												
		Seq primer: M13 FORWARD												
		POLYA=Yes												
		Location/Qualifiers												
		1..696												
		/organism="Homo sapiens"												
		/mol_type="mRNA"												
		/db_xref="taxon:9606"												
		/clone="UI-H-DHI-aww-h-01-0-U"												
		/tissue_type="Metastatic Chondrosarcoma"												
		/dev_stage="Adult"												
		/lab_host="DHI0B (Life Technologies)"												
		/note="Organ: Lung; Vector: pRTT3-Pac (Pharmacia) with a												
		modified polylinker; Site 1: EcoR I; Site 2: Not I;												
		NCI CGAP_DHI is a normalized cDNA library containing the												
		following tissue(s): VS-8 Cell line from Metastatic												
		Chondrosarcoma in Lung. The library was constructed												
		according to Bonaldo, Lennon and Soares, Genome Research,												
		6:791-806, 1996. First strand cDNA synthesis was primed												
		with an oligo-dT primer containing a Not I site. Double												
		stranded cDNA was ligated to an EcoR I adaptor, digested	</											

Qy	1046	GC	CGAGAGTGGAAAGGAGCATTAAGAAATCCCATTTGGAAGTACAGAAAGAAAGCTCC	1105
Db	539	GC	CGAGAGTGGAAAGGAGCATTAAGAAATCCCATTTGGAAGTACAGAAAGAAAGCTCC	480
Qy	1106	TG	TTAAAAAGCTGCTCTCTGAATTTTCAAAAGTCAAAACAAATCTCTAAGCACTCTCC	1165
Db	479	TG	TTAAAAAGCTGCTCTCTGAATTTTCAAAAGTCAAAACAAATCTCTAAGCACTCTCC	420
Qy	1166	TAT	TAAAGAGAAACCCGTGTTCTTATATCGAAACGTGTTGTAACGTGAATTGAGAG	1225
Db	419	TAT	TAAAGAGAAACCCGTGTTCTTATATCGAAACGTGTTGTAACGTGAATTGAGAG	360
Qy	1226	CCA	AGAAACCCCAAGAAAGCCCGGTCTTCAGTGAACCCCAACCCCAAGACTCTCACTCC	1285
Db	359	CCA	AGAAACCCCAAGAAAGCCCGGTCTTCAGTGAACCCCAACCCCAAGACTCTCACTCC	300
Qy	1286	CCA	AAAGGAGCCAGACCCCATCCCAAGAGAAAGCCTTCTCAAGTGAATGAAGATT	1345
Db	299	CCA	AAAGGAGCCAGACCCCATCCCAAGAGAAAGCCTTCTCAAGTGAATGAAGATT	240
Qy	1346	GCG	TACCTTTCACCAAGAAATGATTTGTGTGTGTCAGACACGAGCTCCCATACC	1405
Db	239	GCG	TACCTTTCACCAAGAAATGATTTGTGTGTGTCAGACACGAGCTCCCATACC	180
Qy	1406	GTT	ACCTTACGGGAATCTCTCCCAAGAGAGAGACTGTAGCAAG	1453
Db	179	GTT	ACCTTACGGGAATCTCTCCCAAGAGAGAGACTGTAGCAAG	132
RESULT 57				
BM661885/c				
LOCUS	527 bp	MRNA	linear	EST 27-FEB-2002
DEFINITION	UI-E-CK1-abn-h-08-0-UI.s1 UI-E-CK1 Homo sapiens cDNA clone			
ACCESSION	BM661885			
VERSION	BM661885.1	GI:18965987		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 527)			
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.			
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery			
JOURNAL	Genome Res. 6 (9), 791-806 (1996)			
MEDLINE	97044477			
PUBMED	8889548			
COMMENT	Contact: Soares, MB Coordinated Laboratory for Computational Genomics University of Iowa 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: bento-soares@uiowa.edu Tissue Procurement: Dr. Gregg Hageman CDNA Library preparation: Dr. M. Bento Soares, University of Iowa CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). Seq primer: M13 Forward POLYA=yes.			
FEATURES				
Source	Location/Qualifiers			
	1..527			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="UI-E-CK1-abn-h-08-0-UI"			
	/tissue_type="Retina Foveal and Macular"			
	/dev_stage="adult"			
	/lab_host="DH10B (Life Technologies) (T1 phage resistant)"			

QY	4301	TTATTAAGAGTTCAATTGTAATTAACATTTGGCTGTAAATAA	4345
Db	45	TTATTAAGAGTTCAATTGTAATTAACATTTGGCTGTAAATAA	1
RESULT 59			
LOCUS	AM173363/c		
DEFINITION	AM173363	524 bp	mRNA
ACCESSION	xj78e12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone		linear EST 16-NOV-1999
VERSION	AM173363		
KEYWORDS	AM173363.1 GI:6439311		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 524)		
JOURNAL	NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap .		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
FEATURES	Unpublished (1997)		
SOURCE	Contact: Robert Strausberg, Ph.D.		
	Email: cga@bbs-remail.nih.gov		
	This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.		
	Seq primer: -40UP from Gldco		
	High quality sequence stop: 465.		
	Location/Qualifiers		
	1..524		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2663374"		
	/lab_host="DH10B"		
	/clone_lib="Soares_NFL_T_GBC_S1"		
	/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The diverter was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."		
ORIGIN			
Query Match	12.0%; Score 524; DB 2; Length 524;		
Best Local Similarity	100.0%; Pred. No. 3.1e-193;		
Matches 524;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	3821	AGTTTCTGACTTGAAGTGAAGTGAAGATCCAGAGTGTCTATCTGGCCAGATTAAAG	38880
Db	524	AGTTTCTGACTTGAAGTGAAGTGAAGATCCAGAGTGTCTATCTGGCCAGATTAAAG	465
QY	3881	TAGATTCTATTTCTCTGGTCTCTCCCTCCCTGAGAGACTTTATTTATTTGTCCTCT	39440
Db	464	TAGATTCTATTTCTCTGGTCTCTCCCTCCCTGAGAGACTTTATTTATTTGTCCTCT	405
QY	3941	TCTAGGTTAATTTCTCTTGAATTTGAATTTGATTTGAAGAAGAGTGTGACAGATTAGC	40000
Db	404	TCTAGGTTAATTTCTCTTGAATTTGAATTTGATTTGAAGAAGAGTGTGACAGATTAGC	345
QY	4001	AAAGTTCCAAAGTGCATAATTAACGTGTGTGAAGTGTGGGGGAAAATTAATCTTATTT	40600
Db	344	AAAGTTCCAAAGTGCATAATTAACGTGTGTGAAGTGTGGGGGAAAATTAATCTTATTT	285
QY	4061	TCCCTACATGGGATACAACTGTGAATTTCAATCTTCAACGAAGGCCCTGCAATTCGC	41200
Db	284	TCCCTACATGGGATACAACTGTGAATTTCAATCTTCAACGAAGGCCCTGCAATTCGC	225

Oy		4121	TAAACATGATGTGGTCTTTTCCTTTACCAAGATTGAAGTAGTGTAATAAATTAATAA	4180
Dd		224	TAAACATGATGTGGTCTTTTCCTTTACCAAGATTGAAGTAGTGTAATAAATTAATAA	165
Oy		4181	AAATGCTGTGCTGTACTTCCAGCTTTGTTTTATGCCCATTTGCATATGTGTGTGTGT	4240
Dd		164	AAATGCTGTGCTGTACTTCCAGCTTTGTTTTATGCCCATTTGCATATGTGTGTGTGT	105
Oy		4241	TGTAATTCATTAACCTTTTGATACCAATTTTCGATGTGTAAATTTGGTGTCTTGTAATATC	4300
Dd		104	TGTAATTCATTAACCTTTTGATACCAATTTTCGATGTGTAAATTTGGTGTCTTGTAATATC	45
Oy		4301	TTAATAAGAGTCATGTGAATTAATAACTATTGTGGCTGTAAAA	4344
Dd		44	TTAATAAGAGTCATGTGAATTAATAACTATTGTGGCTGTAAAA	1
RESULT 60				
BM711612				
LOCUS				
DEFINITION		BM711612	529 bp mRNA linear EST 28-FEB-2002	
ACCESSION		U1-E-CL1-atf-m-11-0-U1.r1 U1-E-CL1 Homo sapiens cDNA clone		
VERSION		U1-E-CL1-atf-m-11-0-U1.5, mRNA sequence.		
KEYWORDS		BM711612		
SOURCE		BM711612.1 GI:19024870		
ORGANISM		EST.		
REFERENCE		Homo sapiens (human)		
AUTHORS		Homo sapiens		
TITLE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 529) Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery Genome Res. 6 (9), 791-806 (1996)		
JOURNAL		Genome Res. 6 (9), 791-806 (1996)		
MEDLINE		97044477		
PUBMED		8889548		
COMMENT		Contact: Soares, MB Coordinated Laboratory for Computational Genomics University of Iowa 375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA Tel.: 319 335 8250 Fax: 319 335 9565 Email: Bento-soares@uiowa.edu Tissue Procurement: Dr. Gregg Hageman CDNA Library preparation: Dr. M. Bento Soares, University of Iowa CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). Seq primer: M13 Reverse.		
FEATURES				
source				
		Location/Qualifiers		
		1..529		
		/organism="Homo sapiens"		
		/mol_type="mRNA"		
		/db_xref="taxon:9606"		
		/clone="U1-E-CL1-atf-m-11-0-U1"		
		/tissue_type="human retina"		
		/dev_stage="adult"		
		/lab_host="DH10B (Life Technologies) (T1 phage resistant)"		
		/note="Organ: eye; Vector: pRT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; U1-E-CL1 is a normalized cDNA library containing the following tissue(s): retina. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The		

ORIGIN

sequence tag for this library is CCGCG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

Query Match 12.0%; Score 524; DB 4; Length 529;
Best Local Similarity 100.0%; Pred. No. 3.1e-193;
Matches 524; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3835 AGTAGCTGAGAGAAATCCACGAGTGTCTATCTGCGCAGATTTAAGTAGATTCTATTTCC 3894
DB 1 AGTAGCTGAGAGAAATCCACGAGTGTCTATCTGCGCAGATTTAAGTAGATTCTATTTCC 60
QY 3895 TTGGTTCCTCCCTCCCTGAGGACCTTTATTTATTTGTCCTCTTTAGTTAATTC 3954
DB 61 TTGGTTCCTCCCTCCCTGAGGACCTTTATTTATTTGTCCTCTTTAGTTAATTC 120
QY 3955 CCTTGATTTGACTTTGTTGAGAGAGGTTGACAGTAGATAGCAAAAGTTCCAAAGTGC 4014
DB 121 CCTTGATTTGACTTTGTTGAGAGAGGTTGACAGTAGATAGCAAAAGTTCCAAAGTGC 180
QY 4015 AAAATTAACAGTGTGTAGAGTGTGGGGGAAAAATAGTCTTATTTTCCCTACATGGGAT 4074
DB 181 AAAATTAACAGTGTGTAGAGTGTGGGGGAAAAATAGTCTTATTTTCCCTACATGGGAT 240
QY 4075 ACAACACTGTGAATTAATCTTCACTGAAGGCCCTGCAAGTTCTCTAAACATAGTTGT 4134
DB 241 ACAACACTGTGAATTAATCTTCACTGAAGGCCCTGCAAGTTCTCTAAACATAGTTGT 300
QY 4135 TTGTTTTCTTTAACAAGTTAAGCTAGTGTAAATTAATAAAAAATTTGCTGTGCG 4194
DB 301 TTGTTTTCTTTAACAAGTTAAGCTAGTGTAAATTAATAAAAAATTTGCTGTGCG 360
QY 4195 TCTACTTACGCTTTGTTTATAGCCCATTTCAATGTGTGTGTGTATTAATCAACT 4254
DB 361 TCTACTTACGCTTTGTTTATAGCCCATTTCAATGTGTGTGTGTATTAATCAACT 420
QY 4255 TTGATACCATTTCTGATGTGTAAATTTGTTGTTTAAATATCTTTAAGAGTTCA 4314
DB 421 TTGATACCATTTCTGATGTGTAAATTTGTTGTTTAAATATCTTTAAGAGTTCA 480
QY 4315 ATTGTAAATTAACATTTGTTGCTGTAAATAAAAAATTTGCTGTGCG 4358
DB 481 ATTGTAAATTAACATTTGTTGCTGTAAATAAAAAATTTGCTGTGCG 524

RESULT 61 533 bp mRNA linear EST 27-FEB-2002
BM65523/c UI-E-CL1-afB-m-11-0-UI.s1 UI-E-CL1 Homo sapiens cDNA clone
LOCUS UI-E-CL1-afB-m-11-0-UI.3, mRNA sequence.
DEFINITION
ACCESSION BM65523
VERSION BM65523.1 GI:18972490
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 533)
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548

COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman

cDNA library preparation by: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Forward
POLYA=yes.

FEATURES

source

Location/Qualifiers
1..533
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CL1-afB-m-11-0-UI"
/tissue_type="human retina"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-CL1"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CL1 is a normalized cDNA library containing the following tissue(s): retina. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGCG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
TAG_TISSUE=human retina
TAG_LIB=UI-E-CL1
TAG_SEQ=CCGCG"

ORIGIN

Query Match 12.0%; Score 524; DB 4; Length 533;
Best Local Similarity 100.0%; Pred. No. 3.1e-193;
Matches 524; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3835 AGTAGCTGAGAGAAATCCACGAGTGTCTATCTGCGCAGATTTAAGTAGATTCTATTTCC 3894
DB 524 AGTAGCTGAGAGAAATCCACGAGTGTCTATCTGCGCAGATTTAAGTAGATTCTATTTCC 465
QY 3895 TTGGTTCCTCCCTCCCTGAGGACCTTTATTTATTTGTCCTCTTTAGTTAATTC 3954
DB 464 TTGGTTCCTCCCTCCCTGAGGACCTTTATTTATTTGTCCTCTTTAGTTAATTC 405
QY 3955 CCTTGATTTGACTTTGTTGAGAGAGGTTGACAGTAGATAGCAAAAGTTCCAAAGTGC 4014
DB 404 CCTTGATTTGACTTTGTTGAGAGAGGTTGACAGTAGATAGCAAAAGTTCCAAAGTGC 345
QY 4015 AAAATTAACAGTGTGTAGAGTGTGGGGGAAAAATAGTCTTATTTTCCCTACATGGGAT 4074
DB 344 AAAATTAACAGTGTGTAGAGTGTGGGGGAAAAATAGTCTTATTTTCCCTACATGGGAT 285
QY 4075 ACAACACTGTGAATTAATCTTCACTGAAGGCCCTGCAAGTTCTCTAAACATAGTTGT 4134
DB 284 ACAACACTGTGAATTAATCTTCACTGAAGGCCCTGCAAGTTCTCTAAACATAGTTGT 225
QY 4135 TTGTTTTCTTTAACAAGTTAAGCTAGTGTAAATTAATAAAAAATTTGCTGTGCG 4194
DB 224 TTGTTTTCTTTAACAAGTTAAGCTAGTGTAAATTAATAAAAAATTTGCTGTGCG 165
QY 4195 TCTACTTACGCTTTGTTTATAGCCCATTTCAATGTGTGTGTGTATTAATCAACT 4254
DB 164 TCTACTTACGCTTTGTTTATAGCCCATTTCAATGTGTGTGTGTATTAATCAACT 105
QY 4255 TTGATACCATTTCTGATGTGTAAATTTGTTGTTTAAATATCTTTAAGAGTTCA 4314
DB 104 TTGATACCATTTCTGATGTGTAAATTTGTTGTTTAAATATCTTTAAGAGTTCA 45

QY	4315	ATTGTAATAATAACTATGTGGCTGTTAAAAA	43158
Db	44	ATTGTAATAATAACTATGTGGCTGTTAAAAA	1
RESULT 62			
LOCUS	BM264268	541 bp	RNA linear EST 18-DEC-2001
DEFINITION	ig33b05.y1 HR85 islet Homo sapiens	CDNA 5', mRNA sequence.	
ACCESSION	BM264268		
VERSION	BM264268.1	GI:17927308	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Bufo variegatus, Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. (bases 1 to 541)		
REFERENCE	Leinhardt, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kastner, K., Hillier, L., Marra, M., Page, D., Wylie, J., Martin, J., Bliscan, A., Schmitt, A., Theising, B., Ralder, E., Konko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, V., Williams, T., Jackson, Y., and Bowers, Y.		
AUTHORS	Endocrine Pancreas Consortium		
TITLE	Unpublished (2000)		
JOURNAL	Other ESTs: ig33b05.x1		
COMMENT	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue Endocrine Pancreas Consortium Harvard University, Howard Hughes Medical Institute Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138 Tel: 617-495-1812 Fax: 617-495-8557 Email: dmelton@mol.harvard.edu Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@lm.wustl.edu) Seg primer: -40RP from Glibco High quality sequence stop: 469. Location/Qualifiers 1. 541 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /tissue_type="Purified pancreatic islet" /lab_host="RDH10B" /clone_lib="HR85 islet" /note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1: NotI; Site 2: XhoI; CDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permut Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."		
FEATURES			
SOURCE			
ORIGIN			
Query Match	12.0%	Score 523;	DB 4; Length 541;
Best Local Similarity	100.0%;	Pred. No. 7,7e-193;	
Matches 523;	Conservative	Mismatches 0;	Indels 0; Gaps 0;
QY	292	CCGGTGGGGGGGCGAAGGGCCGGGGCTGTACTCCCGCGGGCGGGCCCCCGGCGACG	351
Db	541	CCGGTGGGGGGGCGAAGGGCCGGGGCTGTACTCCCGCGGGCGGGCCCCCGGCGACG	482
QY	352	AGGAAGAAGCTGGGGCGGCTTCGGTGCCTTGGCCCTGTCGGCCCCCGGCGACGAAGCAAG	411
Db	481	AGGAAGAAGCTGGGGCGGCTTCGGTGCCTTGGCCCTGTCGGCCCCCGGCGACGAAGCAAG	422
QY	412	CCGGCAATGGGGGGGAGCTTCGGCCGAGCCGAGCCGGAGTGCAGCCCCCGGCGCCCAATGC	471

Db 421 CCGGCAATTGGAGGGGAAGCTTCGCCAGACCGAAGCCGAGTCAAGCCCCCGGCACCAGATATC 362

Oy 472 AGGCGGTGCTGCTGCCATTTCAGA-CGGGCTCTCTTCGTGGGGGGGGCCAAAGACTTA-GGCGCT 531

Db 361 AGGCGGTGCTGCTGCCATTTCAGA-CGGGCTCTCTTCGTGGGGGGGGCCAAAGACTTA-GGCGCT 302

Oy 532 GGGCTGGGGA-CAAAGGTGGGGGGCGGCTCCCCCGCTGGCACCGGCTTGCGAG-CCGGGCGGAG 591

Db 301 GGGCTGGGGA-CAAAGGTGGGGGGCGGCTCCCCCGCTGGCACCGGCTTGCGAG-CCGGGCGGAG 242

Oy 592 CCCCAACCACTACCTCTTCCCGGGGCGGCACCCCTCGGCCAACCCGACCGCCGGAGCC 651

Db 241 CCCCAACCACTACCTCTTCCCGGGGCGGCACCCCTCGGCCAACCCGACCGCCGGAGCC 182

Oy 652 TGGCGGCGCAGCGAGGGCAGATGGAAGATATGAGAAAGAGCCCTCTCGGGGTGTTGACG 711

Db 181 TGGCGGCGCAGCGAGGGCAGATGGAAGAGTATGAGAAAGAGCCCTCTCGGGGTGTTGACG 122

Oy 712 GCTCGGAGAGCTTCAGTCAGGCGGCGGCTCCCTCAAAGCATTCCTTCGCTGAATTGACG 771

Db 121 GCTCGGAGAGCTTCAGTCAGGCGGCGGCTCCCTCAAAGCATTCCTTCGCTGAATTGACG 62

Oy 772 TCATCGAACAAGCAGCAGCAGCAGCGAGCTGCAGGCGAAAGAAAAGA .814

Db 61 TCATCGAACAAGCAGCAGCAGCAGCTGCAGGCGAAAGAAAAGA 19

RESULT 63
LOCUS AW665845/c 522 bp mRNA linear EST 06-APR-2000

DEFINITION h194f04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone

VERSION IMAGE:2979967 3 , mRNA sequence.

KEYWORDS AW665845

SOURCE AW665845.1 GI:7458394

ORGANISM EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-rc@mail.nih.gov

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco

High quality sequence stop: 465.

location/Qualifiers

1..522

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2979967"

/lab_host="DH10B"

/clone_lib="Soares_NFL_T_GBC_S1"

/note="Organ: pooled; Vector: pTRT3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCL CGAP GCBI) were mixed, and as circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
1.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

JOURNAL COMMENT

REFERENCE AUTHORS

TITLE

FEATURES SOURCE

ORIGIN


```

Db      421 AATGAGATTTTGGCGTACCTTTCCACCACAGAAATGATTGTGTGTTGGCCAGCCT 480
Qy      1395 CCCCCATCACCGTTTACCATTCAGGGA 1420
Db      481 CCCCACATCACCGTTTACCATTCAGGGA 506

RESULT 67
BQ711779      985 bp      mRNA      linear      EST 16-JUL-2002
LOCUS      AGENCOURT 8487897 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6300573
DEFINITION      5', mRNA sequence.
ACCESSION      BQ711779
VERSION      BQ711779.1 GI:21850678
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS      1 (bases 1 to 985)
TITLE      NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT      Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: Dr. Mark Watson
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLCM2515 row: a column: 22
High quality sequence stop: 592.
Location/Qualifiers
1.985
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6300573"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_113"
/notes="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC library."

ORIGIN
Query Match      11.5%; Score 502; DB 5; Length 985;
Best Local Similarity 100.0%; Pred.No. 1.1e-184; Mismatches 0; Indels 0; Gaps 0;
Matches 502; Conservative 0;

Qy      569 CACCGCTCGGAGCCCGCGGAGACCCACACCTACTCTCCCGGCGCCACCCCTCGC 628
Db      10 CACCGCTCGGAGCCCGCGGAGACCCACACCTACTCTCCCGGCGCCACCCCTCGC 69
Qy      629 GCCCACCGCACCGCGCGGACCTTGCGCGCGACGAGGAGGAGATGAGAGATGAGAA 688
Db      70 GCCCACCGCACCGCGCGGACCTTGCGCGCGACGAGGAGGAGATGAGAGATGAGAA 129
Qy      689 GAGCCCTCTCGGGGGGTGCTGGCGCTCGGAGAGCTCCAGTCAGGCGCCCTGCTCAACA 748
Db      130 GAGCCCTCTCGGGGGGTGCTGGCGCTCGGAGAGCTCCAGTCAGGCGCCCTGCTCAACA 189
Qy      749 GATCTTCTGCTCAATTGAGACCTCATCGAAGACAGACAGACAGAGCTGCGCAAGA 808
Db      190 GATCTTCTGCTCAATTGAGACCTCATCGAAGACAGACAGACAGAGCTGCGCAAGA 249

```

```

Qy      809 AAAGGATCGAGGAGCTGAAGTCAGAGAGACACCGCTCTTGCTCGATTGAACGTAT 868
Db      250 AAAGGATCGAGGAGCTGAAGTCAGAGAGACACCGCTCTTGCTCGATTGAAGTAT 309
Qy      869 GGAAGGCGGATGCACTGTGTAAGAGAGATACGAGAAAGAGACACAAGCTGTTTCA 928
Db      310 GGAAGGCGGATGCACTGTGTAAGAGAGATACGAGAAAGAGACACAAGCTGTTTCA 369
Qy      929 GGGCTATGAACTGAAGAGAGAGAGAAACAGAGCTATCTGAGAAATTTAACTGAGT 988
Db      370 GGGCTATGAACTGAAGAGAGAGAGAAACAGAGCTATCTGAGAAATTTAACTGAGT 429
Qy      989 CCAGCCGAGCTTTCCGAGACATCCAGACTGCTGCCCAAGCCCTTCAATGTGAGG 1048
Db      430 CCAGCCGAGCTTTCCGAGACATCCAGACTGCTGCCCAAGCCCTTCAATGTGAGG 489
Qy      1049 GAGTGGAAAGGACATTAAGG 1070
Db      490 GAGTGGAAAGGACATTAAGG 511

RESULT 68
BM970469/c      686 bp      mRNA      linear      EST 20-FEB-2003
LOCUS      UI-CF-ECL1-10-0-UI-51
DEFINITION      UI-CF-ECL1-10-0-UI-51 UI-CF-ECL1 Homo sapiens cDNA clone
ACCESSION      BM970469
VERSION      BM970469.1 GI:19588056
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS      1 (bases 1 to 686)
TITLE      Bonaldo,M.F., Lennon,G. and Soares,M.B.
JOURNAL      Normalization and subtraction: two approaches to facilitate gene
MEDLINE      discovery
PUBMED      genome Res. 6 (9), 791-806 (1996)
COMMENT      97044477
Contact: McCray, PB
McCrays Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 1-3', >Poly_A#simple_repeat (matched complement)
Seq primer: M13 FORWARD
POLYA=yes.
Location/Qualifiers
1.686
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-ECL1-10-0-UI-51"
/tissue_type="lung"
/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-ECL1"
/notes="Organ: lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-ECL1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was

```


constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AAGGCTTAC.

TAG TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371 and 380-383
TAG LIB=UI-CF-ECL
TAG_SEQ=AAGGCTTAC"

ORIGIN

Query Match 11.4%; Score 499; DB 5; Length 686;
Best Local Similarity 99.5%; Pred. No. 1.7e-183;
Matches 649; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3535 GGGCTCATCTCAACCTTAAGAGAGATTTCTAGAAAACGGGCGAGATTTCCTT 3594
DB 670 GGGCTCATCTCAACCTTAAGAGAGATTTCTAGAAAACGGGCGAGATTTCCTT 611
QY 3595 GTTCTCATCTTAATATGAGAGAGCTTCACTTCTTACTTACCTATGATAT 3654
DB 610 GTTCTCATCTTAATATGAGAGAGCTTCACTTCTTACTTACCTATGATAT 551
QY 3655 TTCTTGTAAAGTGTCCAAAAGAAAAGAACCCATCAGTGTCTTGAATTGTTCTT 3714
DB 550 TTCTTGTAAAGTGTCCAAAAGAAAAGAACCCATCAGTGTCTTGAATTGTTCTT 491
QY 3715 TGAATCCCTGATTTCTTGAATTTGAGAGAGTGTGGGTTCCAAATTTGGGATAGT 3774
DB 490 TGAATCCCTGATTTCTTGAATTTGAGAGAGTGTGGGTTCCAAATTTGGGATAGT 431
QY 3775 TAGCAATTTAACCATTTGTTTGTGCTTACCAGGAGGAGTCCCGAGTTTCTGACTTGA 3834
DB 430 TAGCAATTTAACCATTTGTTTGTGCTTACCAGGAGGAGTCCCGAGTTTCTGACTTGA 371
QY 3835 AGTAGACTGAAAGATCCACGAGTGTCTATCTGGCCAGATTAGTAGATTCTATTTC 3894
DB 370 AGTAGACTGAAAGATCCACGAGTGTCTATCTGGCCAGATTAGTAGATTCTATTTC 311
QY 3895 TTGGTCTCCCTCCCTCCGAGAGCTTATTTATTTGTCCTCCCTCTAGTTAATTC 3954
DB 310 TTGGTCTCCCTCCCTCCGAGAGCTTATTTATTTGTCCTCCCTCTAGTTAATTC 251
QY 3955 CCTTGTATTTGATTTGTTGAGAGAGTTGACAGTAGATTAGCAAAATTTCCAAATGTC 4014
DB 250 CCTTGTATTTGATTTGTTGAGAGAGTTGACAGTAGATTAGCAAAATTTCCAAATGTC 191
QY 4015 AAAATTAACAGTGTAGAGTGTGGGGGAAATTAAGTCTTATTTTCCCTACATGGAGT 4074
DB 190 AAAATTAACAGTGTAGAGTGTGGGGGAAATTAAGTCTTATTTTCCCTACATGGAGT 131
QY 4075 ACAACACTGTGATTTCACTTCACTGAAAGCCCTGAGTTCCTCTAAACATAGTTGT 4134
DB 130 ACAACACTGTGATTTCACTTCACTGAAAGCCCTGAGTTCCTCTAAACATAGTTGT 71
QY 4135 TTGTTTTCTTTAAAGAGTTTAAAGTGTATTAATTAATTAATTAATTAATTAATTAAT 4186
DB 70 TTGTTTTCTTTAAAGAGTTTAAAGTGTATTAATTAATTAATTAATTAATTAATTAAT 19

RESULT 69
AI800794
LOCUS
DEFINITION w913c04.x1 Soares_NSF_P8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:12364966 3 similar to contains MER22.b3 MER22 repetitive element; mRNA
ACCESSION AI800794
VERSION AI800794.1 GI:536266

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 549)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1197 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 451.
Location/Qualifiers
1. 549
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2364966"
/lab_host="DH10B"
/clone_lib="Soares_NSF_P8_9W_OT_PA_P_S1"
/note="Organ: pooled; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHP pool 1: 309384-310919, 323208-325895 Soares NBHP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NBHP-9W pool 1: 758280-760583, 772104-774407 Soares NBHP pool 1: 304776-306311, 320136-322823, 326280-326653 Soares NBHP pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bernaldo."

ORIGIN

Query Match 11.4%; Score 498; DB 1; Length 549;
Best Local Similarity 100.0%; Pred. No. 4.3e-183;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 TGAATCAAGAGCGGCTGCGGCTGGGCGGCGGCGGAGAGAGCTGGGGCGGCGAG 204
DB 38 TGAATCAAGAGCGGCTGCGGCTGGGCGGCGGCGGAGAGAGCTGGGGCGGCGAG 97
QY 205 CCACTTCTCCCTCCGAGAGCTGAGTCAAGAGCGGCGGCGGCTGCTCTCC 264
DB 98 CCACTTCTCCCTCCGAGAGCTGAGTCAAGAGCGGCGGCGGCTGCTCTCC 157
QY 265 AGGCGGAGAGCGGCGGCTTCCCGGCGGCGGCGGCGGAGAGCGGCGGCTGTAC 324
DB 158 AGGCGGAGAGCGGCGGCTTCCCGGCGGCGGCGGCGGAGAGCGGCGGCTGTAC 217
QY 325 TCCCGCGGCGGCGGCGGCGGCGGAGAGAGAGTGGGCGTTCGTCCTTGC 384
DB 218 TCCCGCGGCGGCGGCGGCGGCGGAGAGAGAGTGGGCGTTCGTCCTTGC 277
QY 278 CTTGTCGCGGCGGCGGCGGAGAGAGAGTGGGCGGAGAGCTGCGGAGCGGAG 444
DB 385 CTTGTCGCGGCGGCGGCGGAGAGAGAGTGGGCGGAGAGCTGCGGAGCGGAG 337
QY 445 CCGGCTGAGAGCGGCGGCGGAGAGTCAAGAGAGTGGGCGGAGAGAGAGAGAGAGAG 504
DB 338 CCGGCTGAGAGCGGCGGCGGAGAGTCAAGAGAGTGGGCGGAGAGAGAGAGAGAGAG 397
QY 505 TGGCGGCGGCGGAG 564

```
Db      398 TGCGCGCGGCAAGAGCCCTAGCGCCTGGGGCAAGGGTGGGGCGCGCTCCCGG 457
Qy      565 CTGCGACCGCTTGAGACCGGGGGGAGACCCCACTACTCTGCGCGGGCGCCACCCG 624
Db      458 CTGCGACCGCTTGAGACCGGGGGGAGACCCCACTACTCTGCGCGGGCGCCACCCG 517
Qy      625 TGCGCGCGGCAAGAGCCCTAGCGCCTGGGGCAAGGGTGGGGCGCGCTCCCGG 642
Db      518 TGCGCGCGGCAAGAGCCCTAGCGCCTGGGGCAAGGGTGGGGCGCGCTCCCGG 535

RESULT 70
Bg178033
LOCUS   Bg178033          951 bp      mRNA      linear      EST 06-FEB-2001
DEFINITION Homo sapiens cDNA clone IMAGE:4429246 5',
            mRNA sequence.
ACCESSION Bg178033
VERSION   Bg178033.1 GI:12684736
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 951)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: DCTD/DTF
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: LHAM10180 row: m column: 23
            High quality sequence stop: 636.
            Location/Qualifiers
                1..951
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4429246"
                /tissue_type="adenocarcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_91"
                /note="Organ: prostate; Vector: pCMV-Sport6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.4 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."

ORIGIN
Query Match      11.4%; Score 497; DB 4; Length 951;
Best Local Similarity 99.8%; Pred. No. 9,7e-183;
Matches 617; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy      2937 TTGTGTAATATCAGAGTAGCACTCCAGAGCACTTGTGTCAGTGAAGATCCCTAT 2996
Db      23  TTGTGTAATATCAGAGTAGCACTCCAGAGCACTTGTGTCAGTGAAGATCCCTAT 82
Qy      2997 TAACTATCTGAAAGAAATAGAGCCAGACCTCTGCTCAATATATAGAAATTCCT 3056
Db      83  TAACTATCTGAAAGAAATAGAGCCAGACCTCTGCTCAATATATAGAAATTCCT 142
Qy      3057 TTCTTTAGTCTTGAAGACTATTTGTGTAAGAAAGTGGGGCTTAATCTCTAGAGGTA 3116
Db      143  TTCTTTAGTCTTGAAGACTATTTGTGTAAGAAAGTGGGGCTTAATCTCTAGAGGTA 202
Qy      3117 GGGGCTTTATCTCTTAAGAGAAATATGTCCCGAGATTAATAGACCTTTAGAGGAGAGC 3176
Db      203  GGGGCTTTATCTCTTAAGAGAAATATGTCCCGAGATTAATAGACCTTTAGAGGAGAGC 262
```

```
Qy      3177 CAAGGTAATGAGGGTGTGTGCTGGCCATCACTGAGAGACGGAAGAGAAATGGATACC 3236
Db      263  CAAGGTAATGAGGGTGTGTGCTGGCCATCACTGAGAGACGGAAGAGAAATGGATACC 322
Qy      3237 ATTGTGGAGAGAGAAAGAAAGTTCCCTCAGGGGCTCCCACTGCTAAAGTTTTTGTGAG 3296
Db      323  ATTGTGGAGAGAGAAAGAAAGTTCCCTCAGGGGCTCCCACTGCTAAAG-TTTTGTGAG 381
Qy      3297 TGTGATCTGTGCTTCTGGAATTTGACTTTTAAAGGAATTAATTTGGCAGCAGCATGTAGT 3356
Db      382  TGTGATCTGTGCTTCTGGAATTTGACTTTTAAAGGAATTAATTTGGCAGCAGCATGTAGT 441
Qy      3357 ATCTTGAGATGATCTTGCTGCTCTTAATTTCTCTTTTGTGTGTGTGTGTGTGTGTG 3416
Db      442  ATCTTGAGATGATCTTGCTGCTCTTAATTTCTCTTTTGTGTGTGTGTGTGTGTGTG 501
Qy      3417 GCTATGGTTTCATTTGTAATCCATCTGCTTAGAGAGTGGGCTCTATTAAGGAAC 3476
Db      502  GCTATGGTTTCATTTGTAATCCATCTGCTTAGAGAGTGGGCTCTATTAAGGAAC 561
Qy      3477 CTGCTGTAACCTTCATTCAGCAGCAAGGATGAGAGAAATAGCACTTAATTCACTAGGG 3536
Db      562  CTGCTGTAACCTTCATTCAGCAGCAAGGATGAGAGAAATAGCACTTAATTCACTAGGG 621
Qy      3537 GCTCTCATCTCACACCTT 3554
Db      622  GCTCTCATCTCACACCTT 639

RESULT 71
Bg167371
LOCUS   Bg167371          551 bp      mRNA      linear      EST 12-SEP-2001
DEFINITION Homo sapiens cDNA clone IMAGE:511571 5',
            mRNA sequence.
ACCESSION Bg167371
VERSION   Bg167371.1 GI:15581604
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 551)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
            Toshiyuki and Piero Carninci (RIKEN)
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: LHAM11789 row: i column: 12
            High quality sequence stop: 551.
            Location/Qualifiers
                1..551
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:511571"
                /tissue_type="hypothalamus"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_96"
                /note="Organ: Brain; Vector: pBluescriptR (modified
                pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
                (gcgcag); Oligo-dT primed using primer
                5'-TTTTTTTTTTTTTTVN-3', size-selected to R05. This is a
                primary library enriched for full-length clones and
```

constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN

```
Query Match 11.3%; Score 495; DB 4; Length 551;
Best Local Similarity 99.8%; Pred. No. 6.3e-182;
Matches 545; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 867 ATGGAAGGCGGAGTGCAGCTGCTGTTAAAGATACGAGAAAGAAAGGACACAGCTGTT 926
DB 6 ATGGAAGGCGGAGTGCAGCTGCTGTTAAAGATACGAGAAAGAAAGGACACAGCTGTT 65
QY 927 CAGGCGTATGAACTGAAAGAGAGAGAGAAACAGAGCTATCTGGAATAATTAACCTGAG 986
DB 66 CAGGCGTATGAACTGAAAGAGAGAGAGAAACAGAGCTATCTGGAATAATTAACCTGAG 125
QY 987 TGCAGCGGAGCTTTCCGAGACATCCAGACTGCTGCTCCAGAGCCCTTCTCATGTGG 1046
DB 126 TGCAGCGGAGCTTTCCGAGACATCCAGACTGCTGCTCCAGAGCCCTTCTCATGTGG 185
QY 1047 CGGAGTGGAAAGGACATTAAGAAATCCCATTTGGAAGTACAGAAAGAAAGACTCT 1106
DB 186 CGGAGTGGAAAGGACATTAAGAAATCCCATTTGGAAGTACAGAAAGAAAGACTCT 245
QY 1107 GTTAAAGGCTGCTGCTGAAATTTTCAAAAGTCAAAACAAAGCTCTTAAGACTCTCT 1166
DB 246 GTTAAAGGCTGCTGCTGAAATTTTCAAAAGTCAAAACAAAGCTCTTAAGACTCTCT 305
QY 1167 ATTAAGAGAGAACCTGCTGCTTCTTATCTGAAACCTGTTGTAACCTGAATTGAGAG 1226
DB 306 ATTAAGAGAGAACCTGCTGCTTCTTATCTGAAACCTGTTGTAACCTGAATTGAGAG 365
QY 1227 CAAGAAACCCGAGAAAGCCCGGCTTTCAGTGGACACCCGACCAAGACTCTCCACTCC 1286
DB 366 CAAGAAACCCGAGAAAGCCCGGCTTTCAGTGGACACCCGACCAAGACTCTCCACTCC 425
QY 1287 CAAAGGAGACCCGAGACCCATCCCAAGAGAAAGCTTCAAGTGAATGAAGATTG 1346
DB 426 CAAAGGAGACCCGAGACCCATCCCAAGAGAAAGCTTCAAGTGAATGAAGATTG 485
QY 1347 CCGTACCTTTCCACAGACAGAAATGATTTGTGTGTCGACACAGCTCCCCCATCAG 1406
DB 486 CCGTACCTTTCCACAGACAGAAATGATTTGTGTGTCGACACAGCTCCCCCATCAG 545
QY 1407 TTACCA 1412
DB 546 TTACCA 551
```

```
RESULT 72
LOCUS B0127983 508 bp mRNA linear EST 15-JUL-2003
DEFINITION 5179g08.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:577847
ACCESSION B0127983
VERSION B0127983.1 GI:20201894
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 508)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,D., Blisstein,A.,
Schmitt,A., Theising,B., Ritter,B., Ronko,I., Bennett,J.,
Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagaratshvili,R.,
Williams,T., Jackson,Y., and Bowers,Y.
TITLE Unpublished (2000)
JOURNAL Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
```

Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: emelton@biochem.harvard.edu

Library was constructed by Dr. J. Ferrer in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center for information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40RP from Gidco
High quality sequence stop: 430.

FEATURES

source

```
1..508
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:577847"
/tissue_type="Insulinoma"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="Human insulinoma"
/note="Organ: pancreas; Vector: pBluescript SK-; Site:1:
XhoI; Site 2: EcoRI; Constructed with lambda ZapIt system
(Stratagene) by Dr. J. Ferrer. In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."
```

ORIGIN

```
Query Match 11.3%; Score 492; DB 5; Length 508;
Best Local Similarity 100.0%; Pred. No. 9.5e-181;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 641 CGCCGGAGCCCTGCGGCGGCGGCGGAGGCGAGATGGAAGATGGAAGAGCCCTCTCGG 700
DB 17 CGCCGGAGCCCTGCGGCGGCGGCGGAGGCGAGATGGAAGATGGAAGAGCCCTCTCGG 76
QY 701 GGGTGTGTGGGCGCTCGGAGGCTCTCAAGTCAAGCGCGCTGCTCAACAGATCTTCTGCT 760
DB 77 GGGTGTGTGGGCGCTCGGAGGCTCTCAAGTCAAGCGCGCTGCTCAACAGATCTTCTGCT 136
QY 761 GCAATTGACCTCATTCAGACAGACAGACAGACAGCTGACAGCCCAAGAAAGAGATCGA 820
DB 137 GCAATTGACCTCATTCAGACAGACAGACAGACAGCTGACAGCCCAAGAAAGAGATCGA 196
QY 821 GAGCTGAAGTCAAGAGAGACACAGCTCTTCTGCTCGGATTTGAAGCTATGGAAGCGGAT 880
DB 197 GAGCTGAAGTCAAGAGAGACACAGCTCTTCTGCTCGGATTTGAAGCTATGGAAGCGGAT 256
QY 881 GCAAGCTGTAAGAGAGATTAAGAGAGAAAGGCAAGCTGTTCAAGGCTATGGAAC 940
DB 257 GCAAGCTGTAAGAGAGATTAAGAGAGAAAGGCAAGCTGTTCAAGGCTATGGAAC 316
QY 941 TGAAGAGAGAGAGAAACAGAGCTATCTGGAATAATTAACCTGAGTGCAGCGGAGCT 1000
DB 317 TGAAGAGAGAGAGAAACAGAGCTATCTGGAATAATTAACCTGAGTGCAGCGGAGCT 376
QY 1001 TTCCGAGACATCCCAAGCTGCTGCTCCCAAGCCCTTCTCATGTGGCGGAGTGAAGGG 1060
DB 377 TTCCGAGACATCCCAAGCTGCTGCTCCCAAGCCCTTCTCATGTGGCGGAGTGAAGGG 436
QY 1061 ACATTAAGAGAAATCCCATTTTGAAGTACAGAAAGAAAGACTCTCTTAAGAAAGCTGGC 1120
DB 437 ACATTAAGAGAAATCCCATTTTGAAGTACAGAAAGAAAGACTCTCTTAAGAAAGCTGGC 496
QY 1121 TCCTGAATTTTC 1132
```

Db	497	TCCTGATTTTC	508
RESULT 73			
LOCUS	W03542		
DEFINITION	W03542	510 bp	mrna
ACCESSION	Y599B09.f1	Soares melanocyte 2Nbhm	linear
VERSION	IMAGE:291161	5', mRNA sequence.	EST 19-APR-1996
KEYWORDS	W03542		
SOURCE	W03542.1	GI:1275387	
ORGANISM	EST.		
	Homo sapiens (human)		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 510)		
AUTHORS	Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maira,M., Parsons,D., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.		
TITLE	The Washu-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK		

```
FEATURES      Location/Qualifiers
source        1. .510
```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="CDB:3884291"
/db_xref="taxon:9606"
/clone="IMAGE:291161"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares melanocyte 2NbHM"
/notes="vector: pRTT3D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGAGCGCCGACATTTTCTTTTCTTTT 3'],
TGGTCC-Stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRTT3 vector
(Pharmacia). Library constructed by Bento Soares and
M. Felina Bonaido. RNA from normal foreskin melanocytes
(FBS374) was kindly provided by Dr. Anthony P. Albino."

```

ORIGIN

	Query Match	11.3%	Score 492;	DB 7;	Length 510;
	Best Local Similarity	100.0%	Pred. No. 9.5e-191;		
	Matches 492;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	3560	GGAGATTCTGAAAACTGGGCCAGATTTCTTGTTCTCCATCATTTTAATGCGCAG	3615		
Db	15	GGAGATTCTGAAAACTGGGCCAGATTTCTTGTTCTCCATCATTTTAATGCGCAG	74		
QY	3620	GCTGTTCAAGTTTCTTACTCTTACTCTAATGTATATTTCTTGTAACGTGTCAAAAAGAA	3677		
Db	75	GCTGTTCAAGTTTCTTACTCTTACTCTAATGTATATTTCTTGTAACGTGTCAAAAAGAA	134		
QY	3680	AAAGACCACCATCAAGTGTCTTGAGTTTGTTCTTGATCCCTCAGTTTCTCTTGATTT	373		
Db	135	AAAGACCACCATCAAGTGTCTTGAGTTTGTTCTTGATCCCTCAGTTTCTCTTGATTT	194		
QY	3740	CAGCATGTGTCGGGTTCTCTAATTTTGGGTAATGATTAGCAAAATTTAACCATGTGTTGT	3799		

Db	195	CAGATGTGTGCGGTTCTTAATTTTGGTATNAGATTGACAAATTAACTCATGTGTGT	254
Qy	3800	GCCCTACCCAGGGGACCTCCCAAGTTTCTGACTTGAAGTAGACTGAGAAATCCACGAG	3859
Db	255	GCCCTACCCAGGGGACCTCCCAAGTTTCTGACTGAAATGACTGAGAAATCCACGAG	314
Qy	3860	TGCTATCTGGCCAGATTAAAGTAGATTCFAATTTCTTGTTCTCCCTCCCTGAGGACC	3919
Db	315	TGCTATCTGGCCAGATTAAAGTAGATTCFAATTTCTTGTTCTCCCTCCCTGAGGACC	374
Qy	3920	TCTTATTTTATTTGTCCTCTCTTCTAGGTTAATTCCTTGATTTGAATTGTTGAGAG	3979
Db	375	TCTTATTTTATTTGTCCTCTCTTCTAGGTTAATTCCTTGATTTGAATTGTTGAGAG	434
Qy	3880	GAGTTTGACAGTAGATTAGCAAAATTCGCAAGTTCGCAAAATTAACGCTGTGTAAGGTG	4039
Db	435	GAGTTTGACAGTAGATTAGCAAAATTCGCAAGTTCGCAAAATTAACGCTGTGTAAGGTG	494
Qy	4040	GGGGAATAATTAG 4051	
Db	495	GGGGAATAATTAG 506	

RESULT	74			
LOCUS	AA746655/c			
DEFINITION	AA746655	491 bp	mRNA	linear EST 16-JAN-1998
ACCSSION	U27810.1	NCI	CGAP	G4
VERSION	AA746655			
KEYWORDS	AA746655.1	GI:2786641		
SOURCE	EST			
ORGANISM	Homo sapiens			
	Homo sapiens (human)			
	Homo sapiens			

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 491)	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap .		
	National Cancer Institute,	Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index			
	unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			

Email: CGAPbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E Consortium/BLNT at:
www-bio.linn.gov/bdbp/image/image.html
Seq primer: -40m13 fwd, Et from Amersham
High quality sequence stop: 471.

FEATURES

SOURCE

```

ORIGIN
Query Match      11.2%; Score 491; DB 1; Length 491;

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1257331"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCT CGAP GC4"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73D
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Ronaldo. "

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2005, 05:01:04 ; Search time 2109 Seconds
(without alignments)
12271.744 Million cell updates/sec

Title: US-10-054-935-1
Perfect score: 4372
Sequence: 1 cagtcctcgcaccccccgcac.....aaaaaaaaaaaaaa 4372

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4390206 seqs, 2959870667 residues

Word size : 15

Total number of hits satisfying chosen parameters: 376363

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

- 1: N_Geneseq_16Dec04:*
- 2: geneseq1980s:*
- 3: geneseq1990s:*
- 4: geneseq2000s:*
- 5: geneseq2001s:*
- 6: geneseq2002s:*
- 7: geneseq2003s:*
- 8: geneseq2003as:*
- 9: geneseq2003bs:*
- 10: geneseq2003cs:*
- 11: geneseq2003ds:*
- 12: geneseq2004as:*
- 13: geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4372	100.0	4372	8 ACC72044	ACC72044 BCUI041 g
2	2888	66.1	3425	12 AD064204	Adq64204 Novel hum
3	1370	31.3	1755	12 AD086569	Adq86569 Human tum
4	1370	31.3	1755	13 AD084001	Adq84001 Human tum
5	1370	31.3	1755	13 AD083399	Adq83399 Human tum
6	1260	28.8	1509	2 AA24879	AA24879 Human sec
7	1260	28.8	1509	8 ADA39700	ADA39700 Human sec
8	1260	28.8	1509	8 ACC50379	ACC50379 Human sec
9	1260	28.8	1509	9 ADH91097	ADH91097 Human sec
10	1260	28.8	1509	10 ADA55894	ADA55894 Gene enco
11	903	20.7	1545	11 AD130841	AD130841 Human cDN
12	837	19.1	1529	3 AAF18193	AAF18193 Lung can
13	732	16.7	2887	3 AAC76874	AAC76874 Human ORF
14	487	11.1	1290	11 ADI30996	ADI30996 Human cDN
15	469	10.7	910	4 AA193984	AA193984 Human neu
16	449	10.3	904	4 AA193983	AA193983 Human neu
17	427	9.8	893	4 AA194300	AA194300 Human neu
18	422	9.7	1143	10 ADF57688	ADF57688 Human pol
19	379	8.7	2887	3 AAC76874	AAC76874 Human ORF
20	357	8.2	365	9 ACH42852	ACH42852 Human foe

21	306	7.0	903	4 AA194299	AA194299 Human neu
22	284	6.5	451	2 AAT22113	AAT22113 Human gen
23	270	6.2	2106	10 ADF59738	ADF59738 Human con
24	230	5.3	230	6 ABV88816	ABV88816 Human col
25	213	4.9	411	9 ACH46937	ACH46937 Human inf
26	152	3.5	542	6 ABK39516	ABK39516 DNA enco
27	152	3.5	542	8 ACA11845	ACA11845 Human lun
28	152	3.5	542	10 ADH47073	ADH47073 Human lun
29	152	3.5	542	13 ADJ20992	ADJ20992 Human lun
30	146	3.3	865	12 ADP84445	ADP84445 Human bre
31	134	3.1	211	5 ABV56446	ABV56446 Human pro
32	110	2.5	615	6 ABK62818	ABK62818 Rat seque
33	106	2.4	317	3 AAC25273	AAC25273 Human sec
34	101	2.3	482	8 ACA03031	ACA03031 Lung can
35	63	1.4	346	6 ABX52399	ABX52399 Bovine ES
36	60	1.4	60	6 ABN49572	ABN49572 Human spl
37	51	1.2	590	6 ABK62109	ABK62109 Rat seque
38	51	1.2	590	10 ADB55014	ADB55014 toxicity-
39	49	1.1	196	4 AAH36707	AAH36707 Human col
40	47	1.1	50	3 AAC11178	AAC11178 Human sec
41	38	0.9	827	2 AAT47669	AAT47669 Human g p
42	38	0.9	827	6 ABX14564	ABX14564 Human gam
43	38	0.9	1654	3 AAF18137	AAF18137 Lung can
44	37	0.8	241	8 ABX47809	ABX47809 Bovine ES
45	37	0.8	243	5 ABV56639	ABV56639 Human pro
46	37	0.8	403	4 AA183077	AA183077 Human pol
47	37	0.8	410	4 AA183503	AA183503 Human pol
48	37	0.8	410	9 ACH21614	ACH21614 Human adu
49	37	0.8	418	9 ACH19590	ACH19590 Human adu
50	37	0.8	432	4 AA187359	AA187359 Human pol
51	37	0.8	446	4 AAL08202	AAL08202 Human bre
52	37	0.8	450	9 ACH23288	ACH23288 Human adu
53	37	0.8	458	2 AAQ90525	AAQ90525 Rat SIII
54	37	0.8	467	4 AAT59993	AAT59993 Rat RNA p
55	37	0.8	467	4 AAL19112	AAL19112 Human bre
56	37	0.8	632	8 ACA61000	ACA61000 Novel hum
57	37	0.8	632	9 ADB47644	ADB47644 Human cDN
58	37	0.8	632	13 ADRA40538	ADRA40538 Human ORP
59	37	0.8	647	8 ACA04606	ACA04606 cDNA enco
60	37	0.8	674	11 ACN87141	ACN87141 Breast ca
61	37	0.8	782	4 AAH33250	AAH33250 Human col
62	37	0.8	1218	2 AAT48447	AAT48447 Human TGF
63	37	0.8	1227	2 AAT48446	AAT48446 Human TGF
64	37	0.8	1228	2 AAT48449	AAT48449 Human TGF
65	37	0.8	1345	8 AAD51362	AAD51362 Soybean f
66	37	0.8	1441	2 AAT48444	AAT48444 Human TGF
67	37	0.8	1663	3 AAC59302	AAC59302 Human sec
68	37	0.8	1663	8 ADB273679	ADB273679 Secreted
69	37	0.8	1663	8 ADA98154	ADA98154 Human sec
70	37	0.8	1663	10 ABT16860	ABT16860 Human sec
71	37	0.8	1663	10 ABZ67274	ABZ67274 Human sec
72	37	0.8	1665	13 ADP24257	ADP24257 PRO polyp
73	37	0.8	1703	3 AAC96942	AAC96942 Human sec
74	37	0.8	1703	6 ABL90376	ABL90376 Human pol
75	37	0.8	1706	8 AAL53985	AAL53985 cDNA enco
76	37	0.8	1883	2 AAU59670	AAU59670 Human sec
77	37	0.8	1883	3 ABE73657	ABE73657 Human cDN
78	37	0.8	1883	9 ACD82800	ACD82800 cDNA sequ
79	37	0.8	1883	10 AD122885	AD122885 cDNA enco
80	37	0.8	1883	12 ADH73887	ADH73887 Human sec
81	37	0.8	2030	9 AAL62910	AAL62910 Wheat cDN
82	37	0.8	2044	12 ADN04035	ADN04035 Antipor
83	37	0.8	2439	12 ADL82952	ADL82952 Human NF-
84	37	0.8	2439	13 ADR14070	ADR14070 Human NF-
85	37	0.8	2522	5 ABV29355	ABV29355 Human pro
86	37	0.8	2522	5 ABV21123	ABV21123 Human pro
87	37	0.8	2522	5 ABV23498	ABV23498 Human pro
88	37	0.8	2522	5 ABV26965	ABV26965 Human pro
89	37	0.8	2602	12 ADM67085	ADM67085 Human NF-
90	37	0.8	2807	13 ADR14518	ADR14518 Human NF-
91	37	0.8	2808	11 ADN39683	ADN39683 Cancer/an
92	37	0.8	2915	3 AAA16668	AAA16668 Human sec
93	37	0.8	3034	4 AAS03899	AAS03899 Human sec

94	37	0.8	3034	8	AB733661	Secred
95	37	0.8	3034	8	ADA98146	Human sec
96	37	0.8	3034	10	ABT16854	Human sec
97	37	0.8	3034	10	AB672255	Human sec
98	37	0.8	3366	4	AA246484	Nucleotid
99	37	0.8	3366	4	AA24706	Nucleotid
100	37	0.8	3517	10	ADG76377	Human Inc

ALIGNMENTS

RESULT 1
 ID ACC72044 standard; DNA; 4372 BP.
 XX
 AC ACC72044;
 DT 08-JUL-2003 (first entry)
 XX
 DE BC11041 gene #SEQ ID 65.
 XX
 KW Breast cancer; cytostatic; gene therapy; antisense therapy; regulated;
 KM drug discovery; clinical medicine; forensic medicine; gene;
 KW chromosome 7q21.1; ds.
 XX
 OS Homo sapiens.
 PN W02003029421-A2.
 PD 10-APR-2003.
 PF 02-OCT-2002; 2002WO-US031287.
 XX
 PR 03-OCT-2001; 2001US-0326526P.
 PR 14-MAY-2002; 2002US-00144194.
 XX
 PA (ORIG-) ORIGENE TECHNOLOGIES INC.
 XX
 P1 Sun Z, Li X, Fan W, Kovacs KF, Jay G;
 XX
 DR WPI; 2003-381623/36.
 DR P-PSDB; ABR58310.
 XX
 PT New isolated human differentially-regulated breast cancer polynucleotide
 PT and polypeptide, useful for diagnosing, staging, prognosticating,
 PT preventing and/or treating diseases and conditions relating to breast
 PT cancer.
 XX
 PS Claim 2; SEQ ID NO 65; 127bp + Sequence Listing; English.
 XX
 CC The invention relates to isolated polynucleotides which are
 CC differentially-regulated in breast cancer. The methods and compositions
 CC of the present invention are useful for detecting, diagnosing, staging,
 CC monitoring, prognosticating, preventing and/or treating diseases and
 CC conditions relating to breast cancer, and may be used in gene therapy or
 CC antisense therapy. They can also be used in research, drug discovery,
 CC clinical medicine and forensic medicine. Sequences given in records
 CC ACC72012-ACG72017 represent polynucleotides of the invention that are
 CC differentially-regulated in breast cancer. NOTE: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SO Sequence 4372 BP; 1136 A; 1012 C; 1077 G; 1147 T; 0 U; 0 Other;

Query Match 100.0%; Score 4372; DB 8; Length 4372;
 Best Local Similarity 100.0%; Pct. No. 0;
 Matches 4372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CAGGCTCTGACAGCCCCCGACCTCGCCCTTCCCAACCCCTCTCCGCTCGGATCCCGG 60
 1 CAGGCTCTGACAGCCCCCGACCTCGCCCTTCCCAACCCCTCTCCGCTCGGATCCCGG 60

Qy	61	CGCTGCTCCGGAACAATAATGACATGAGATCCGCGGTGTTCAAGGCGGCGCGCCCTG	120
Db	61	CGCTGCTCCGGAACAATAATGACATGAGATCCGCGGTGTTCAAGGCGGCGCGCCCTG	120
Qy	121	CGGCGGCAATCTGTAGACAGCACTGAGCTGACGCGGCTGCGGCTGCGGCTCGG	180
Db	121	CGGCGGCAATCTGTAGACAGCACTGAGCTGACGCGGCTGCGGCTGCGGCTCGG	180
Qy	121	CGGCGGCAATCTGTAGACAGCACTGAGCTGACGCGGCTGCGGCTGCGGCTCGG	180
Db	121	CGGCGGCAATCTGTAGACAGCACTGAGCTGACGCGGCTGCGGCTGCGGCTCGG	180
Qy	181	AGGACGAGCCTGAGGCGGCGCGGAGCCCACTTCTCCCGGCGACCGTAACCTAAGAGC	240
Db	181	AGGACGAGCCTGAGGCGGCGCGGAGCCCACTTCTCCCGGCGACCGTAACCTAAGAGC	240
Qy	241	CGGAGGCCCGCGTGGGCTCTCTCCAGAGGCGGAGCCCGCGGCTTCCCGGCGCTGCG	300
Db	241	CGGAGGCCCGCGTGGGCTCTCTCCAGAGGCGGAGCCCGCGGCTTCCCGGCGCTGCG	300
Qy	301	GCGGCAAGGCGCGGCGCTTGTATCTCCCGGCGGCGGCGGCGGCGGCGGCGGAGAGA	360
Db	301	GCGGCAAGGCGCGGCGCTTGTATCTCCCGGCGGCGGCGGCGGCGGCGGCGGAGAGA	360
Qy	361	GCTGGGCGGTTGAGTGCCCTTGCCTTGCCTGATCCGCGCGCGGCGGCGGCGGCACTTG	420
Db	361	GCTGGGCGGTTGAGTGCCCTTGCCTTGCCTGATCCGCGCGCGGCGGCGGCGGCACTTG	420
Qy	421	GCGGCGGCTGCTGCGGCGGAGCGGCTGCAAGCCCGGCGGCGGCGGCGGCGGCGGCTG	480
Db	421	GCGGCGGCTGCTGCGGCGGAGCGGCTGCAAGCCCGGCGGCGGCGGCGGCGGCGGCTG	480
Qy	481	TGCGCATTCAGAGGCGGCTCTGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	540
Db	481	TGCGCATTCAGAGGCGGCTCTGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	540
Qy	541	ACAAGGCTGGGCGGCGCTCCCGGCTGCGACCGGCTGCGACCGGCGGAGCCCGCAC	600
Db	541	ACAAGGCTGGGCGGCGCTCCCGGCTGCGACCGGCTGCGACCGGCGGAGCCCGCAC	600
Qy	601	TACTCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	660
Db	601	TACTCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	660
Qy	661	GCGAGGCGAGATGGAAGATGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	720
Db	661	GCGAGGCGAGATGGAAGATGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	720
Qy	721	CTTCCAGTCAAGCGCGCTGCTCAACAGATCCTTGTGCTCAATTGCACTCATGAGC	780
Db	721	CTTCCAGTCAAGCGCGCTGCTCAACAGATCCTTGTGCTCAATTGCACTCATGAGC	780
Qy	781	AGGAGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAG	840
Db	781	AGGAGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAG	840
Qy	841	ACAAGGCTCTTGTGCTGAGTTGAAGTATGAGAAAGCGGAGTGAAGTGTAAAGAGATA	900
Db	841	ACAAGGCTCTTGTGCTGAGTTGAAGTATGAGAAAGCGGAGTGAAGTGTAAAGAGATA	900
Qy	901	ACGAGAAAGAAAGGCGACACAGCTGTTTCAAGGCTATGAAACTGAAGAGAGAGAAACAG	960
Db	901	ACGAGAAAGAAAGGCGACACAGCTGTTTCAAGGCTATGAAACTGAAGAGAGAGAAACAG	960
Qy	961	AGCATCTGAGAAATTTAACTGAGAGTGCAGCGCGGAGCTTTCGAGACATCCAGACTC	1020
Db	961	AGCATCTGAGAAATTTAACTGAGAGTGCAGCGCGGAGCTTTCGAGACATCCAGACTC	1020
Qy	1021	TGCTCTCCAGCCCTTCTCATGTGGGCGGAGTGAAGGAGCATTTAAAGAAATCCCAT	1080
Db	1021	TGCTCTCCAGCCCTTCTCATGTGGGCGGAGTGAAGGAGCATTTAAAGAAATCCCAT	1080
Qy	1081	TTGGAAGTAAAGAAAGAAAGATCTCTGTTAAAGAGTGGCTCTCGTAATTTTCAAAAGTCA	1140
Db	1081	TTGGAAGTAAAGAAAGAAAGATCTCTGTTAAAGAGTGGCTCTCGTAATTTTCAAAAGTCA	1140

QY 1141 AAAAACAACCTCTTAAGCACTCTCTTAATTAAGAGAACCTGTGTCTTATCTGAAA 1200
DB 1141 AAAAACAACCTCTTAAGCACTCTCTTAATTAAGAGAACCTGTGTCTTATCTGAAA 1200
QY 1201 CTGTGTGTAAAGCTGAATTAAGAGAACCAAGAAACCCCAAGAAAGCCCGGTCTTCAGTGG 1260
DB 1201 CTGTGTGTAAAGCTGAATTAAGAGAACCAAGAAACCCCAAGAAAGCCCGGTCTTCAGTGG 1260
QY 1261 AACACCCCAAGAAAGCTCTCCACTCCCAAAAGGGAACCCAGACCCATCCCAAGAGAAAG 1320
DB 1261 AACACCCCAAGAAAGCTCTCCACTCCCAAAAGGGAACCCAGACCCATCCCAAGAGAAAG 1320
QY 1321 CCTTCTCAAGTGAATTAAGAAATTTGCGGTACCTTTCCACACAGAAATGATTTGTGTC 1380
DB 1321 CCTTCTCAAGTGAATTAAGAAATTTGCGGTACCTTTCCACACAGAAATGATTTGTGTC 1380
QY 1381 GTTGGACACCGCTCCCGCTACCGTTACCATTAAGGGAATCTCTCCAAAGAGAGG 1440
DB 1381 GTTGGACACCGCTCCCGCTACCGTTACCATTAAGGGAATCTCTCCAAAGAGAGG 1440
QY 1441 AGACTGAGCAAGGTGTCTGATGCATCAAGTGTGAGAGAACTTCAGTCTGGCTG 1500
DB 1441 AGACTGAGCAAGGTGTCTGATGCATCAAGTGTGAGAGAACTTCAGTCTGGCTG 1500
QY 1501 TTCTCTTCTGAGAGGACCACTCAGTAGAGCTCTAAGGGAACCAATCTTCAGACCTTT 1560
DB 1501 TTCTCTTCTGAGAGGACCACTCAGTAGAGCTCTAAGGGAACCAATCTTCAGACCTTT 1560
QY 1561 TGGAGAACCTGATGACAGTGTGTTTTCGAAGCGGCATGCAAACTGAGACTGAGTGA 1620
DB 1561 TGGAGAACCTGATGACAGTGTGTTTTCGAAGCGGCATGCAAACTGAGACTGAGTGA 1620
QY 1621 AGGAGAAAGAAATGAGGATATTCAGAGATCAGGGAACAAAGATTTTACAGGACTGC 1680
DB 1621 AGGAGAAAGAAATGAGGATATTCAGAGATCAGGGAACAAAGATTTTACAGGACTGC 1680
QY 1681 AGCTCAAGATGTATTAAGAAAGAAAGAAATTCAGAGATTCAGGCTGAGGTTTACTCATTTT 1740
DB 1681 AGCTCAAGATGTATTAAGAAAGAAAGAAATTCAGAGATTCAGGCTGAGGTTTACTCATTTT 1740
QY 1741 TCCCTGAGCCAGATGATGTTGAAAGTTGATGATTAACCCCTTCTGCTGTGTGATCAT 1800
DB 1741 TCCCTGAGCCAGATGATGTTGAAAGTTGATGATTAACCCCTTCTGCTGTGTGATCAT 1800
QY 1801 TTGAGAGACCATTAACCAAAATTAATCCACAGAAATTTGAGCTTACCTGTGTGATGAGC 1860
DB 1801 TTGAGAGACCATTAACCAAAATTAATCCACAGAAATTTGAGCTTACCTGTGTGATGAGC 1860
QY 1861 GTAGCCGATCAGATTTGAGATTCAGAAAGAGCAACCTCACCGGACGTGTGAGAAAT 1920
DB 1861 GTAGCCGATCAGATTTGAGATTCAGAAAGAGCAACCTCACCGGACGTGTGAGAAAT 1920
QY 1921 AGCTGTGCTGCAAGAAACCTGTCTTCAGATGATTTGAGATGTCATGAGAGTGGC 1980
DB 1921 AGCTGTGCTGCAAGAAACCTGTCTTCAGATGATTTGAGATGTCATGAGAGTGGC 1980
QY 1981 AGAGACCTGTATATGTAACCTGTGTCTCAGATATGTTATCACTCGCTGATATACCTT 2040
DB 1981 AGAGACCTGTATATGTAACCTGTGTCTCAGATATGTTATCACTCGCTGATATACCTT 2040
QY 2041 TCATACTCTCTGACTTGTGTTTTCATTAATCTGATTTTCAAAAACCTTTTCAATTCGGCT 2100
DB 2041 TCATACTCTCTGACTTGTGTTTTCATTAATCTGATTTTCAAAAACCTTTTCAATTCGGCT 2100
QY 2101 AATTTGATGTTATGAGAGGTGATTTGGGATTTCTTTTCTTTTGGGAAATGGGCTCTC 2160
DB 2101 AATTTGATGTTATGAGAGGTGATTTGGGATTTCTTTTCTTTTGGGAAATGGGCTCTC 2160
QY 2161 AAGCTTAAGCTATATGATGAGAGTTCAGAAATTTGAGGAGGTCTGTTTCTATACATTTGC 2220
DB 2161 AAGCTTAAGCTATATGATGAGAGTTCAGAAATTTGAGGAGGTCTGTTTCTATACATTTGC 2220
QY 2221 CTATGTTAAAGGGGTAAAGGGGCTCTCTTATTAAGATGAGAGATGAAGACGCCCT 2280

DB 2221 CTATGTTAAAGGGGTAAAGGGGCTCTCTTATTAAGATGAGAGATGAAGAGGCCCT 2280
QY 2281 TCCCTTAAGAGCTGTGCTGATGAGCACTCTTCACCTGTGTACACCTCTTATAGTGG 2340
DB 2281 TCCCTTAAGAGCTGTGCTGATGAGCACTCTTCACCTGTGTACACCTCTTATAGTGG 2340
QY 2341 GTATAGTGAATTTTAAACCTTAAATTAAGCAAACTCACCATGAGCTTTAGAGACAG 2400
DB 2341 GTATAGTGAATTTTAAACCTTAAATTAAGCAAACTCACCATGAGCTTTAGAGACAG 2400
QY 2401 AAGAGATGACAGTGAAGCGATGAAGCAAGCATCTTCACAGATGAGAAAGACATCG 2460
DB 2401 AAGAGATGACAGTGAAGCGATGAAGCAAGCATCTTCACAGATGAGAAAGACATCG 2460
QY 2461 GAGAGTTGGATGATTAAGTGTCTGAAAAGATGATGTTTCAATTAAGCTATTCGTGTAC 2520
DB 2461 GAGAGTTGGATGATGATTAAGTGTCTGAAAAGATGATGTTTCAATTAAGCTATTCGTGTAC 2520
QY 2521 AGTCATGTGGGAAGGATGTTTGGCTGTGATTAATTTTTCAGTTAAATGATTAACAATTC 2580
DB 2521 AGTCATGTGGGAAGGATGTTTGGCTGTGATTAATTTTTCAGTTAAATGATTAACAATTC 2580
QY 2581 TTTTACTGCTCAAAAACCAAAATCTTTGAAAAGAAAGTGGGATGTTTACAGACA 2640
DB 2581 TTTTACTGCTCAAAAACCAAAATCTTTGAAAAGAAAGTGGGATGTTTACAGACA 2640
QY 2641 AGTTAAGCTGTAAACAAAGGACCTTAATTTGGATGAGCAAGCCAAACCTGTATAA 2700
DB 2641 AGTTAAGCTGTAAACAAAGGACCTTAATTTGGATGAGCAAGCCAAACCTGTATAA 2700
QY 2701 TGTCTTGTATCATCATCATCTTCAGATATCTCTCATTTGGGCTTACCTTTTACAGCA 2760
DB 2701 TGTCTTGTATCATCATCATCTTCAGATATCTCTCATTTGGGCTTACCTTTTACAGCA 2760
QY 2761 ACTCTTGTGTGAGATGAGACTTAAGAGGATGAGGAGAGTGTGAATATAGTGT 2820
DB 2761 ACTCTTGTGTGAGATGAGACTTAAGAGGATGAGGAGAGTGTGAATATAGTGT 2820
QY 2821 TCCCTTGGCTGGCAAAATGTCTATCATCTTGAACAAACAGATGTATAGACTTCTC 2880
DB 2821 TCCCTTGGCTGGCAAAATGTCTATCATCTTGAACAAACAGATGTATAGACTTCTC 2880
QY 2881 ATTCACTTGTAAATTAATTTGATGTATGATGATGATGATGATGATGATGATGATGATGAT 2940
DB 2881 ATTCACTTGTAAATTAATTTGATGTATGATGATGATGATGATGATGATGATGATGATGAT 2940
QY 2941 TTAATAATCAGATGAGCACTCCAGGCACTTTGTCTCAGTGTATGATGATGATGATGAT 3000
DB 2941 TTAATAATCAGATGAGCACTCCAGGCACTTTGTCTCAGTGTATGATGATGATGATGAT 3000
QY 3001 TATCTGAAGGAAATATGAGGCAAGACCTTGTCTCAATATATATGAGATGATGATGATGAT 3060
DB 3001 TATCTGAAGGAAATATGAGGCAAGACCTTGTCTCAATATATATGAGATGATGATGATGAT 3060
QY 3061 TTAGCTTCAAGACTATGTTGTGAAAACAAAGTATGAGGAGTATCTCTTAAGAGTATGAGG 3120
DB 3061 TTAGCTTCAAGACTATGTTGTGAAAACAAAGTATGAGGAGTATCTCTTAAGAGTATGAGG 3120
QY 3121 CTTTATCTCTTAAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180
DB 3121 CTTTATCTCTTAAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180
QY 3181 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
DB 3181 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
QY 3241 TGGGAG 3300
DB 3241 TGGGAG 3300
QY 3301 GATCTGTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3360

```
Db 3301 GATCTGCTCTCTCGGATTTGACTTTTAAAGAAATTATTCGGCAGACATGTAATTC 3360
Qy 3361 TTGGATGATCTGCTCTCTTATTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGT 3420
Db 3361 TTGGATGATCTGCTCTCTTATTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGT 3420
Qy 3421 TGGGTTTATTTGTAATCTCCATCTGCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
Db 3421 TGGGTTTATTTGTAATCTCCATCTGCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
Qy 3481 TGTAAACCTTCATTCGACGAAGATGTAGAGAAATAGCAATTAATTCACCTAGGGGCTC 3540
Db 3481 TGTAAACCTTCATTCGACGAAGATGTAGAGAAATAGCAATTAATTCACCTAGGGGCTC 3540
Qy 3541 TCATCTCACACCTTAAAGAGAGAGATTTCTAAGAAAACCTGGGCCAGATTTCTTTGTTCTC 3600
Db 3541 TCATCTCACACCTTAAAGAGAGATTTCTAAGAAAACCTGGGCCAGATTTCTTTGTTCTC 3600
Qy 3601 CATCATTTTAATGTGGCAGAGCTGTTCAGTTTCTTACTCTTACCTATGTATATTTCTTC 3660
Db 3601 CATCATTTTAATGTGGCAGAGCTGTTCAGTTTCTTACTCTTACCTATGTATATTTCTTC 3660
Qy 3661 GTAACTGTCCAAAAAGAAAAGAACCCCAATCAGTGTCTCTTGACTTTGTTCTTTGATCC 3720
Db 3661 GTAACTGTCCAAAAAGAAAAGAACCCCAATCAGTGTCTCTTGACTTTGTTCTTTGATCC 3720
Qy 3721 CTCAGTTTCTTCTTGAATTCAGCATGTGTGCGGTTCTCTAATTTTGGGTATGATGACAA 3780
Db 3721 CTCAGTTTCTTCTTGAATTCAGCATGTGTGCGGTTCTCTAATTTTGGGTATGATGACAA 3780
Qy 3781 ATTTAACCAATGTGTTTGTGCCCCCCTACCCAGGGACCTCCAGTTTCTGAATTAAGTAA 3840
Db 3781 ATTTAACCAATGTGTTTGTGCCCCCCTACCCAGGGACCTCCAGTTTCTGAATTAAGTAA 3840
Qy 3841 CTGAGAGAGATCCAGAGGTCTATCTGCGCAGATTTAATGATTAATTTCTTCTGTT 3900
Db 3841 CTGAGAGAGATCCAGAGGTCTATCTGCGCAGATTTAATGATTAATTTCTTCTGTT 3900
Qy 3901 CTCCTCTCTCCCTGAGGACCTCTTATTTATTTATTTCCCTCTCTTGAAGTTATTTCTCT 3960
Db 3901 CTCCTCTCTCCCTGAGGACCTCTTATTTATTTATTTCCCTCTCTTGAAGTTATTTCTCT 3960
Qy 3961 ATTTGACTTTGTGAGAGAGAGGTGTGACAGTATTAATTAATTAATTAATTAATTAAT 4020
Db 3961 ATTTGACTTTGTGAGAGAGAGGTGTGACAGTATTAATTAATTAATTAATTAATTAAT 4020
Qy 4021 ACAGTGTGTTAGAGTGTGAGGAGGAGAAATAGTCTTATTTTCCCTACATGAGGATACACA 4080
Db 4021 ACAGTGTGTTAGAGTGTGAGGAGGAGAAATAGTCTTATTTTCCCTACATGAGGATACACA 4080
Qy 4081 CTGTGAATTCATCTTCAACTGAAGGCGCTGCAAGTTCTCTTAAACAATTTGTTGTTT 4140
Db 4081 CTGTGAATTCATCTTCAACTGAAGGCGCTGCAAGTTCTCTTAAACAATTTGTTGTTT 4140
Qy 4141 TTCTTTAAACAAGTTTAAAGTATGTTAATTAATTAATTAATTAATTAATTAATTAATTA 4200
Db 4141 TTCTTTAAACAAGTTTAAAGTATGTTAATTAATTAATTAATTAATTAATTAATTAATTA 4200
Qy 4201 TTAGCTTTGTTTATGAGGCAATTCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4260
Db 4201 TTAGCTTTGTTTATGAGGCAATTCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4260
Qy 4261 ACCATTTCTGATGTGTAATTTGTTGTCTTGTAAATATCTTAAAGATTCATTAATTTGA 4320
Db 4261 ACCATTTCTGATGTGTAATTTGTTGTCTTGTAAATATCTTAAAGATTCATTAATTTGA 4320
Qy 4321 AATTAACCTATTTGCTGTTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAA 4372
Db 4321 AATTAACCTATTTGCTGTTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAA 4372
Qy 4372 AATTAACCTATTTGCTGTTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAA 4372
Db 4372 AATTAACCTATTTGCTGTTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAA 4372
```

RESULT 2
AD064204

```
ID AD064204 standard; cDNA; 3425 BP.
XX
AC AD064204;
XX
DT 07-OCT-2004 (first entry)
XX
DE Novel human cDNA sequence #1365.
XX
KW ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
KW cytostatic; gene therapy; diagnostic marker; morbid state; osteoporosis;
KW neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW cancer.
XX
OS Homo sapiens.
XX
PN BP1440981-A2.
XX
PD 28-JUL-2004.
XX
PF 21-JUN-2004; 2004EP-00001196.
XX
PR 21-JAN-2003; 2003JP-00102206.
PR 09-MAY-2003; 2003JP-00131392.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Nagai K, Irie R;
XX
DR MPI; 2004-535376/52.
XX
PT P-PDB; AD066392.
XX
PS Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
XX Claim 1; SEQ ID NO 1365; 2449pp; English.
XX
CC The invention relates to 2495 novel polynucleotides (I) and their encoded
CC polypeptides, sequences hybridizing to these nucleotides, sequences
CC encoding partial polypeptides and sequences having 70% or 90% identity to
CC the nucleotide and protein sequences. The nucleotides and polypeptides
CC are useful as diagnostic markers or therapeutic target for the diseases
CC or morbid states. They are also useful for treating osteoporosis,
CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC dementia and various cancers. This sequence corresponds to a nucleotide
CC sequence of the invention.
XX
SQ Sequence 3425 BP; 922 A; 636 C; 763 G; 1104 T; 0 U; 0 Other;
XX
Query Match 66.1%; Score 2888; DB 12; Length 3425;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2888; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1453 GGTGTTGATGTCATCAATGTTGTCAGAGAGAACTTCACTTGGTGTCTTCTTGA 1512
Db 538 GGTGTTGATGTCATCAATGTTGTCAGAGAGAACTTCACTTGGTGTCTTCTTGA 597
Qy 1513 GGGACCACTCAGTAGAGCTCTTAAGGAGCCAAATCTTCAAGCTTTTGAAGAACTGG 1572
Db 598 GGGACCACTCAGTAGAGCTCTTAAGGAGCCAAATCTTCAAGCTTTTGAAGAACTGG 657
Qy 1573 ATGACAGTGTGTTTGAAGCGGCATGCAAACTGAGCTGATGAGAAAGAGAGAAA 1632
Db 658 ATGACAGTGTGTTTGAAGCGGCATGCAAACTGAGCTGATGAGAAAGAGAGAAA 717
Qy 1633 GATGGATATTCAGAGATCAAGGAAACAAAGATTTTAAGGACCTGCTAGAAATG 1692
Db 718 GATGGATATTCAGAGATCAAGGAAACAAAGATTTTAAGGACCTGCTAGAAATG 777
Qy 1693 ATAAAAAGAAAGAAATTCAGAGAACTGAGCTGAGGTTACTATTTTCCCTGAGCCAG 1752
Db 778 ATAAAAAGAAAGAAATTCAGAGAACTGAGCTGAGGTTACTATTTTCCCTGAGCCAG 837
```

QY 1753 ATGATGTGAAGTTGATGATTAACCCCTTCTTGCTGTGTGACATTTGAGACACAT 1812
| | | | |
DB 838 ATGATGTGAAGTTGATGATTAACCCCTTCTTGCTGTGTGACATTTGAGACACAT 897
| | | | |
QY 1813 TACCAAAATTAATCCACAGAAATTTTGAAGTACCCCTGTGTGATGACATTTGAGACACAT 1872
| | | | |
DB 898 TACCAAAATTAATCCACAGAAATTTTGAAGTACCCCTGTGTGATGACATTTGAGACACAT 957
| | | | |
QY 1873 GATTGAGATCCAGAGAGACAAACACCTCACCGGACGTGTGAGAAATAGCTGTGAC 1932
| | | | |
DB 958 GATTGAGATCCAGAGAGACAAACACCTCACCGGACGTGTGAGAAATAGCTGTGAC 1017
| | | | |
QY 1933 AAGAACCTGTCTTCAGATAGTTGACATGCCATTTCCGAGAGTGGCAGAGACCTGTAT 1992
| | | | |
DB 1018 AAGAACCTGTCTTCAGATAGTTGACATGCCATTTCCGAGAGTGGCAGAGACCTGTAT 1077
| | | | |
QY 1993 ATGTGACCTTTGTCTTCACATATGTATCTAGCTGTATTAATACCTTTCACTTCTT 2052
| | | | |
DB 1078 ATGTGACCTTTGTCTTCACATATGTATCTAGCTGTATTAATACCTTTCACTTCTT 1137
| | | | |
QY 2053 GACTTGTCTTCACTTCTGATTTACAAAAAACCTTTTCACTTGTGCTTAATGTGAGTTA 2112
| | | | |
DB 1138 GACTTGTCTTCACTTCTGATTTACAAAAAACCTTTTCACTTGTGCTTAATGTGAGTTA 1197
| | | | |
QY 2113 TGGAGGATGATGGGATTTCTTTCCCTTTTGGGAAATGGGCTCAAGCTAAAGCTA 2172
| | | | |
DB 1198 TGGAGGATGATGGGATTTCTTTCCCTTTTGGGAAATGGGCTCAAGCTAAAGCTA 1257
| | | | |
QY 2173 TAGGATGGCAGATTCAGAAATTTTCAGGGGTCTGTTCTATACATTTGCTTAAAGG 2232
| | | | |
DB 1258 TAGGATGGCAGATTCAGAAATTTTCAGGGGTCTGTTCTATACATTTGCTTAAAGG 1317
| | | | |
QY 2233 GGTAAAGGGCTCTCTTCACTTGAACATGTGGAAGATGAGAGGCCCTTCTTGAAGCT 2292
| | | | |
DB 1318 GGTAAAGGGCTCTCTTCACTTGAACATGTGGAAGATGAGAGGCCCTTCTTGAAGCT 1377
| | | | |
QY 2293 GTGCTGATGAGCAGCTTCTTCAACCTGTGACACCTCTTATATGAGGATATGATGAT 2352
| | | | |
DB 1378 GTGCTGATGAGCAGCTTCTTCAACCTGTGACACCTCTTATATGAGGATATGATGAT 1437
| | | | |
QY 2353 TTAACCTTAATAATAACAAACAACTCACATGAGCTTTTGAAGCAGAAAGAGATGAC 2412
| | | | |
DB 1438 TTAACCTTAATAATAACAAACAACTCACATGAGCTTTTGAAGCAGAAAGAGATGAC 1497
| | | | |
QY 2413 AAGGAGAGATGAGAGACATCTTCAAGAGTAAAGAAAGATGGAGAGTGTGATG 2472
| | | | |
DB 1498 AAGGAGAGATGAGAGACATCTTCAAGAGTAAAGAAAGATGGAGAGTGTGATG 1557
| | | | |
QY 2473 ATTAAGTCTGAAGAGATGATGTTCACTTGAACATATCTGTGATACAGTCACTGGGA 2532
| | | | |
DB 1558 ATTAAGTCTGAAGAGATGATGTTCACTTGAACATATCTGTGATACAGTCACTGGGA 1617
| | | | |
QY 2533 AGGAGATTTGGCTGTGATTAATTTTCACTTGAATGATTAACAAATTTCTTACTGCTCA 2592
| | | | |
DB 1618 AGGAGATTTGGCTGTGATTAATTTTCACTTGAATGATTAACAAATTTCTTACTGCTCA 1677
| | | | |
QY 2593 AAACCAAAATCTTTGAGAAAGAAAGTGGGATGTTAGTTGACAAACAGTACAGCTGT 2652
| | | | |
DB 1678 AAACCAAAATCTTTGAGAAAGAAAGTGGGATGTTAGTTGACAAACAGTACAGCTGT 1737
| | | | |
QY 2653 AAACCAAAAGCATTAGTATTTGGGATGGCAGTCCAAACCTGTATTAATGTCTGTATC 2712
| | | | |
DB 1738 AAACCAAAAGCATTAGTATTTGGGATGGCAGTCCAAACCTGTATTAATGTCTGTATC 1797
| | | | |
QY 2713 ACATCACTTCAAGTATTCCTTCACTTGGGCTTCACTTGTACAGAACTCTTGTGATG 2772
| | | | |
DB 1798 ACATCACTTCAAGTATTCCTTCACTTGGGCTTCACTTGTACAGAACTCTTGTGATG 1857
| | | | |
QY 2773 GGGATAGAGCTTAGGGAGGGTGGGAGAGTGTGAAATAGTGTCTCTTGGGCTGG 2832
| | | | |
DB 1858 GGGATAGAGCTTAGGGAGGGTGGGAGAGTGTGAAATAGTGTCTCTTGGGCTGG 1917
| | | | |
QY 2833 CAATGTCTATCTTGAACAAACAGATGATTAAGAGCTTCTCATTTCACTTTGTA 2892
| | | | |

DB 1918 CAATGTCTATCTTGAACAAACAGATGATTAAGAGCTTCTCATTTCACTTTGTA 1977
| | | | |
QY 2893 AAATATATTTGATGTATGATACATCTTGTCTCTCCCTCCGCTTTGTTAAATATCAG 2952
| | | | |
DB 1978 AAATATATTTGATGTATGATACATCTTGTCTCTCCCTCCGCTTTGTTAAATATCAG 2037
| | | | |
QY 2953 GATAGACCTCCAGGCACTTTGTGTCACTGTAGATCCCTATTAATCTAGTAAAGGA 3012
| | | | |
DB 2038 GATAGACCTCCAGGCACTTTGTGTCACTGTAGATCCCTATTAATCTAGTAAAGGA 2097
| | | | |
QY 3013 AAATAGACCAAGACCTCTGTCTCAATATATAGGAATGCTTTTCTTATGCTTACAG 3072
| | | | |
DB 2098 AAATAGACCAAGACCTCTGTCTCAATATATAGGAATGCTTTTCTTATGCTTACAG 2157
| | | | |
QY 3073 ACTATGTGTGAAGAAACAGTATGGGCTTATCTCTAGAGATGAGGGCTTTATCTTGA 3132
| | | | |
DB 2158 ACTATGTGTGAAGAAACAGTATGGGCTTATCTCTAGAGATGAGGGCTTTATCTTGA 2217
| | | | |
QY 3133 AAGAGATATGTCCAGATTAATTAAGCATTTTAGAGAGAACCAAGTATGAGGATG 3192
| | | | |
DB 2218 AAGAGATATGTCCAGATTAATTAAGCATTTTAGAGAGAACCAAGTATGAGGATG 2277
| | | | |
QY 3193 TGTGCTGCGCCATCAGTGGAGACAGAGAGATGGATACCATTTGTGGAAAGAGAG 3252
| | | | |
DB 2278 TGTGCTGCGCCATCAGTGGAGACAGAGAGATGGATACCATTTGTGGAAAGAGAG 2317
| | | | |
QY 3253 AAAGTTCCCTCAGAGGCGCTCCCATGCTAAGTTTGTGAGATGTGATCTGTGCTTC 3312
| | | | |
DB 2338 AAAGTTCCCTCAGAGGCGCTCCCATGCTAAGTTTGTGAGATGTGATCTGTGCTTC 2397
| | | | |
QY 3313 CTGATTTGACTTTTAAAGGAATTAATCTTGCGCAGACATGATGATTTCTGATGATCTT 3372
| | | | |
DB 2398 CTGATTTGACTTTTAAAGGAATTAATCTTGCGCAGACATGATGATTTCTGATGATCTT 2457
| | | | |
QY 3373 GCTGCTTATATTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3432
| | | | |
DB 2458 GCTGCTTATATTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2517
| | | | |
QY 3433 TGTATCTCCATCTGCTTGAAGAGTGGGCTCTCTATAGAGGAACCTGCTTAAACCTTCA 3492
| | | | |
DB 2518 TGTATCTCCATCTGCTTGAAGAGTGGGCTCTCTATAGAGGAACCTGCTTAAACCTTCA 2577
| | | | |
QY 3493 TGCAGCAGAGATGAGAGAGAAATPAGACTTAATCCACTAGGGGCTCTCATCTCACACC 3552
| | | | |
DB 2578 TGCAGCAGAGATGAGAGAGAAATPAGACTTAATCCACTAGGGGCTCTCATCTCACACC 2637
| | | | |
QY 3553 TTAAGAGAGATTTCTAGAAAACTGGGCAAGATTTCTTGTCTCTCATCTTAT 3612
| | | | |
DB 2638 TTAAGAGAGATTTCTAGAAAACTGGGCAAGATTTCTTGTCTCTCATCTTAT 2697
| | | | |
QY 3613 GTGCAAGGCTGTTCAGTTTCTTACTCTTACCTATGTATGATTTCTTGTAAGCTGTCCA 3672
| | | | |
DB 2698 GTGCAAGGCTGTTCAGTTTCTTACTCTTACCTATGTATGATTTCTTGTAAGCTGTCCA 2757
| | | | |
QY 3673 AAAAGAAAAAGACCAATAGTGTCTGACTGCTTGTCTTGTATGCCCACTTGTCTTC 3732
| | | | |
DB 2758 AAAAGAAAAAGACCAATAGTGTCTGACTGCTTGTCTTGTATGCCCACTTGTCTTC 2817
| | | | |
QY 3733 TTGATTTCAAGCATGTGTGCGGCTCTTAATTTTGGGATAGATTAGCAAAATTAACATTTG 3792
| | | | |
DB 2818 TTGATTTCAAGCATGTGTGCGGCTCTTAATTTTGGGATAGATTAGCAAAATTAACATTTG 2877
| | | | |
QY 3793 TGTGTTGCGCTTACCAAGGGAATCCCAAGTTTCTGACTTGAAGTATGAGAGATC 3852
| | | | |
DB 2878 TGTGTTGCGCTTACCAAGGGAATCCCAAGTTTCTGAAGTATGAGAGATC 2937
| | | | |
QY 3853 CACGAGGTCTATCTGGCCAGATTTAAGTATGATTTCTTGTGCTCCCTCTCCCT 3912
| | | | |
DB 2938 CACGAGGTCTATCTGGCCAGATTTAAGTATGATTTCTTGTGCTCCCTCTCCCT 2997
| | | | |
QY 3913 GAGAGCTCTTATTTTAATGTGCCCTCTTGAAGTAAATCTCTTGTGATTTGA 3972
| | | | |


```
Db 481 GCTGCCACCGCCTCGGACCCGCGGAGACCCCACTACTCTGCGCGGCGCCACCC 540
Qy 624 CTGCGCCCAACCGCCACCGCGGGACCTTGCGGCCAGCGAGGAGCATGAAAGATATG 683
Db 541 CTGCGCGCCACCGCCACCGCGGGACCTTGCGGCCAGCGAGGAGCATGAAAGATATG 600
Qy 684 AGGAAAGACCTCTCGGCGGATGTCGCGGCTCGGAGCCTTCAGTCAGGCGCGCTGCTC 743
Db 601 AGGAAAGACCTCTCGGCGGATGTCGCGGCTCGGAGCCTTCAGTCAGGCGCGCTGCTC 660
Qy 744 AAACAGATCCTTGTGCTGCAATTGACCTCATGGAACAGAGAGAGAGAGAGAGAGAG 803
Db 661 AAACAGATCCTTGTGCTGCAATTGACCTCATGGAACAGAGAGAGAGAGAGAGAGAG 720
Qy 804 AAGGAAAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 863
Db 721 AAGGAAAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy 864 CGTATGGAAGAGCGGATGACGCTGTTAAAGAGATTAACGAAAGAGAGAGAGAGAG 923
Db 781 CGTATGGAAGAGCGGATGACGCTGTTAAAGAGATTAACGAAAGAGAGAGAGAGAG 840
Qy 924 TTTCAGGCGCTATGAACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 983
Db 841 TTTCAGGCGCTATGAACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Qy 984 GAGTGCAGCGGAGCTTTCGAGAGATCCAGAGCTGCGCCAGAGCCCTCTCAATG 1043
Db 901 GAGTGCAGCGGAGCTTTCGAGAGATCCAGAGCTGCGCCAGAGCCCTCTCAATG 960
Qy 1044 GGGCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1103
Db 961 GGGCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Qy 1104 CCTGTTAAAGAGTGGCTCTGTAATTTTCAAAAGTCAAAAGTCAAAAGTCAAAAG 1163
Db 1021 CCTGTTAAAGAGTGGCTCTGTAATTTTCAAAAGTCAAAAGTCAAAAGTCAAAAG 1080
Qy 1164 CCTATTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1223
Db 1081 CCTATTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Qy 1224 AGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1283
Db 1141 AGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Qy 1284 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1343
Db 1201 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 1344 TTGCGGATCCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1403
Db 1261 TTGCGGATCCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy 1404 CCGTTACATTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1453
Db 1321 CCGTTACATTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1370

RESULT 4
ADQ84001
ID ADQ84001 standard; cDNA; 1755 BP.
XX
AC ADQ84001;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human tumour-associated antigenic target (TAT) cDNA sequence #815.
XX
KW human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
XX cancer; cell proliferative disorder; gene; ss.
OS Homo sapiens.
```

```
XX MO2004060270-A2.
PN
XX
XX 22-JUL-2004.
PD
XX
XX 15-OCT-2003; 2003MO-US029126.
PF
XX
XX 18-OCT-2002; 2002US-0418988P.
PR
XX
XX (GETH ) GENENTECH INC.
PA (WU T D.
PA (ZHOU ) ZHOU Y.
XX
XX WU TD, Zhou Y;
XX
XX WPI; 2004-534300/51.
DR
XX
XX New nucleic acid molecule and encoded polypeptide, for diagnosing,
PT preventing or treating cell proliferative disorders such as cancer.
XX
XX Claim 1; SEQ ID NO 815; 5504gp; English.
PS
XX
XX The present invention describes an isolated tumour-associated antigenic
CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
CC sequence identity to (a)-(c); or (e) a sequence that hybridizes to (a) -
CC (c). Also described: (1) an expression vector comprising the above
CC nucleic acid; (2) a host cell comprising the above expression vector; (3)
CC a process for producing a polypeptide; (4) an isolated polypeptide
CC comprising: (a) an amino acid sequence encoded by any of the above
CC nucleotide sequences; (b) an amino acid sequence encoded by the full-
CC length coding region of the above nucleotide sequences; or (c) a sequence
CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
CC comprising the above polypeptide fused to a heterologous polypeptide; (6)
CC an isolated antibody that binds to the above polypeptide; (7) a process
CC for producing the antibody; (8) an isolated oligopeptide that binds to
CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)
CC binding organic molecule that binds to the above polypeptide; (10) a
CC composition of matter comprising the above (chimeric) polypeptide,
CC antibody, oligopeptide or TAT binding organic molecule, in combination
CC with a carrier; (11) an article of manufacture comprising a container and
CC the composition of matter contained within the container; (12) methods of
CC inhibiting the growth of a cell that expresses the above protein, where
CC the growth of the cell is at least in part dependent upon a growth
CC potentiating effect of the above protein; (13) a method of
CC therapeutically treating a mammal having a cancerous tumour comprising
CC cells that express the above protein; (14) a method of determining the
CC presence of a protein in a sample suspected of containing the protein
CC described above; (15) methods of diagnosing the presence of a tumour in a
CC mammal; (16) a method for treating or preventing a cell proliferative
CC disorder associated with increased expression or activity of the above
CC protein; and (17) a method of binding an antibody, oligopeptide or
CC organic molecule to a cell that expresses the protein described above.
CC The TAT sequences have cytostatic activities, and can be used in gene
CC therapy. The composition and methods are useful for diagnosing,
CC preventing or treating cancer. The composition is also used for preparing
CC a medicament for the therapeutic treatment or diagnostic detection of a
CC cell proliferative disorder or cancer. The present sequence represents a
CC human TAT cDNA sequence from the present invention.
XX
XX Sequence 1755 BP; 435 A; 502 C; 505 G; 313 T; 0 U; 0 Other;
SQ
XX
XX Query Match 31.3%; Score 1370; DB 13; Length 1755;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

D 61 CTGGACTAGAGGAGGCTGGGGCTGGGGGCGCCGAGGACGAGCTGGGGGCGCCGA 120
Q 204 GCCCACTTCTCTCCCGGCGACCGTAAGCTCAGAGAGCGGGGCGCCCGCTGCTCTCTC 263
D 121 GCCCACTTCTCTCCCGGCGACCGTAAGCTCAGAGAGCGGGGCGCCCGCTGCTCTCTC 180
Q 264 CAGGGGCGGAGGCGCGGCTTCCCGGCGGCTGGGGCGGCAAGGGCGGGGCTTGTA 323
D 181 CAGGGGCGGAGGCGCGGCTTCCCGGCGGCTGGGGCGGCAAGGGCGGGGCTTGTA 240
Q 324 CTCGCGCGGAGGCGCGGCTTCCCGGCGGCTGGGGCGGCAAGGGCGGGGCTTGTA 383
D 241 CTCGCGCGGAGGCGCGGCTTCCCGGCGGCTGGGGCGGCAAGGGCGGGGCTTGTA 300
Q 384 CCTGTCTCG 443
D 301 CCTGTCTCG 360
Q 444 GCGGCGTGAAGCG 503
D 361 GCGGCGTGAAGCG 420
Q 504 GTGGCG 563
D 421 GTGGCG 480
Q 564 GCTGCCACCGCGCTCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
D 481 GCTGCCACCGCGCTCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Q 624 CTCG 683
D 541 CTCG 600
Q 684 AGGAAAGGCGCTCTCGGCGGCTGGGGCGCTCGGCGGCTCGAATCGCGCGCTGCTC 743
D 601 AGGAAAGGCGCTCTCGGCGGCTGGGGCGCTCGGCGGCTCGAATCGCGCGCTGCTC 660
Q 744 AAACAGATCTTCTGCTGCAATTTGAGCTTCATCGAACAGAGACGAGCTCGAGGCC 803
D 661 AAACAGATCTTCTGCTGCAATTTGAGCTTCATCGAACAGAGACGAGCTCGAGGCC 720
Q 804 AAGGAAAGGAGATCGAGAGCTGAAGTCAAGAGAGACGCTCTCTGCTCGAATTGAA 863
D 721 AAGGAAAGGAGATCGAGAGCTGAAGTCAAGAGAGACGCTCTCTGCTCGAATTGAA 780
Q 864 CGATGAGAAAGCGAGATGAGCTGTAAAGAGATTAACAGAGAAAGGCAAGCTG 923
D 781 CGATGAGAAAGCGAGATGAGCTGTAAAGAGATTAACAGAGAAAGGCAAGCTG 840
Q 924 TTTCAGGGCTATGAACTGAAAGAGAGAGAGAAACAGAGCTATCTGAAATTTAACTG 983
D 841 TTTCAGGGCTATGAACTGAAAGAGAGAGAGAAACAGAGCTATCTGAAATTTAACTG 900
Q 984 GAGTGGCAGCGGAGCTTTCGAGAGATCCCAAGCTGCTCCCAAGCCTTCTCATGT 1043
D 901 GAGTGGCAGCGGAGCTTTCGAGAGATCCCAAGCTGCTCCCAAGCCTTCTCATGT 960
Q 1044 GGGCGGAGTGAAGGAGACATAAAGAGAAATCCCATTTGAAAGTACAGAAAGAAAGACT 1103
D 961 GGGCGGAGTGAAGGAGACATAAAGAGAAATCCCATTTGAAAGTACAGAAAGAAAGACT 1020
Q 1104 CCTGTTAAAAAGCTGCTCTGTAATTTTCAAAAGTCAAAACAACTCTTAAGCACTCT 1163
D 1021 CCTGTTAAAAAGCTGCTCTGTAATTTTCAAAAGTCAAAACAACTCTTAAGCACTCT 1080
Q 1164 CCTATTAAAGAGAGACCTGTGTTCTTATCTGAAAGCTGTTTGAAGCTGAATTGAGG 1223
D 1081 CCTATTAAAGAGAGACCTGTGTTCTTATCTGAAAGCTGTTTGAAGCTGAATTGAGG 1140
Q 1224 AGCCAGAAAGCCCGAGAAAGCCCGGCTTCAGTGGACACCCCAAGCAAGCTCCACT 1283
D 1141 AGCCAGAAAGCCCGAGAAAGCCCGGCTTCAGTGGACACCCCAAGCAAGCTCCACT 1200

Q 1284 CCCCCAAAAGGAGACCCGAGACCCCATCCCAAGAGAAAGCTTCTCAAGTGAATGAAAGAT 1343
D 1201 CCCCCAAAAGGAGACCCGAGACCCCATCCCAAGAGAAAGCTTCTCAAGTGAATGAAAGAT 1260
Q 1344 TTGCGGTACTCTTCCACACAGAAATGTATTGTGTCGTTGGACACAGGCTCCCCATCA 1403
D 1261 TTGCGGTACTCTTCCACACAGAAATGTATTGTGTCGTTGGACACAGGCTCCCCATCA 1320
Q 1404 CCGTTACCATTAACGGAATCTCTCCAAAGAGAGAGAGACTGTAGCAAG 1453
D 1321 CCGTTACCATTAACGGAATCTCTCCAAAGAGAGAGAGACTGTAGCAAG 1370
RESULT 5
AD083399
ID AD083399 standard; cDNA; 1755 BP.
XX
AC AD083399;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human tumour-associated antigenic target (TAT) cDNA sequence #213.
XX
KW human; tumour-associated antigenic target; TAT; cytosolic; gene therapy;
XX cancer; cell proliferative disorder; gene; se.
XX
OS Homo sapiens.
XX
PN MO2004060270-A2.
XX
PD 22-JUL-2004.
XX
PF 15-OCT-2003; 2003WO-US029126.
XX
PR 18-OCT-2002; 2002US-0418988P.
XX
PA (GENT) GENENTECH INC.
XX
PA (WUTD) WU T D.
XX
PA (ZHOU) ZHOU Y.
XX
PI Wu TD, Zhou Y;
XX
XX WPI, 2004-534300/51.
XX
DR New nucleic acid molecule and encoded polypeptide, for diagnosing,
XX preventing or treating cell proliferative disorders such as cancer.
XX
PS Claim 1; SEQ ID NO 213; 5504bp; English.
XX
XX The present invention describes an isolated tumour-associated antigenic
XX target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
XX sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
XX (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
XX sequence identity to (a)-(c); or (e) a sequence that hybridizes to (a)-
XX (c). Also described: (1) an expression vector comprising the above
XX nucleic acid; (2) a host cell comprising the above expression vector; (3)
XX a process for producing a polypeptide; (4) an isolated polypeptide
XX comprising: (a) an amino acid sequence encoded by any of the above
XX nucleotide sequences; (b) an amino acid sequence encoded by the full-
XX length coding region of the above nucleotide sequences; or (c) a sequence
XX having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
XX comprising the above polypeptide fused to a heterologous polypeptide; (6)
XX an isolated antibody that binds to the above polypeptide; (7) a process
XX for producing the antibody; (8) an isolated oligopeptide that binds to
XX the above polypeptide; (9) a tumour-associated antigenic target (TAT)
XX binding organic molecule that binds to the above polypeptide; (10) a
XX composition of matter comprising the above (chimeric) polypeptide,
XX antibody, oligopeptide or TAT binding organic molecule, in combination
XX with a carrier; (11) an article of manufacture comprising a container and
XX the composition of matter contained within the container; (12) methods of
XX inhibiting the growth of a cell that expresses the above protein, where
XX the growth of the cell is at least in part dependent upon a growth

CC potentiating effect of the above protein; (13) a method of
 CC therapeutically treating a mammal having a cancerous tumour comprising
 CC cells that express the above protein; (14) a method of determining the
 CC presence of a protein in a sample suspected of containing the protein
 CC described above; (15) methods of diagnosing the presence of a tumour in a
 CC mammal; (16) a method for treating or preventing a cell proliferative
 CC disorder associated with increased expression or activity of the above
 CC protein; and (17) a method of binding an antibody, oligopeptide or
 CC organic molecule to a cell that expresses the protein described above.
 CC The TAT sequences have cytoskeletal activities, and can be used in gene
 CC therapy. The composition and methods are useful for diagnosing,
 CC preventing or treating cancer. The composition is also used for preparing
 CC a medicament for the therapeutic treatment or diagnostic detection of a
 CC cell proliferative disorder or cancer. The present sequence represents a
 CC human TAT cDNA sequence from the present invention.

xx
 SQ Sequence 1755 BP; 435 A; 502 C; 505 G; 313 T; 0 U; 0 Other;

Query Match 31.3%; Score 1370; DB 13; Length 1755;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 ATGAGATCCCGGTTCCTCAAGCGCGCCGCTCCGCGGCAATCTTGAAGACCGA 143
 DB 1 ATGAGATCCCGGTTCCTCAAGCGCGCCGCTCCGCGGCAATCTTGAAGACCGA 60
 QY 144 CTGAGACTACAGAGCGGCTGCGGCTGCGGCGCGGCGGCGGCGGCGGCGGCGG 203
 DB 61 CTGAGACTACAGAGCGGCTGCGGCTGCGGCGCGGCGGCGGCGGCGGCGGCGG 120
 QY 204 GCCCACTTCTCCCGCGGCAACCTTCAAGAGCGGCGGCGGCGGCGGCGGCGGCGG 263
 DB 121 GCCCACTTCTCCCGCGGCAACCTTCAAGAGCGGCGGCGGCGGCGGCGGCGGCGG 180
 QY 264 CAGGCGCGGAGCGCGCGCTTCCCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCTT 323
 DB 181 CAGGCGCGGAGCGCGCGCTTCCCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCTT 240
 QY 324 CTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 383
 DB 241 CTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 300
 QY 384 CCTGTCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 443
 DB 301 CCTGTCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
 QY 444 GCCGCGTGCAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 503
 DB 361 GCCGCGTGCAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 420
 QY 504 GTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 563
 DB 421 GTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 480
 QY 564 GCTGCGCAGCGGCTGGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 623
 DB 481 GCTGCGCAGCGGCTGGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
 QY 624 CTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 683
 DB 541 CTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 600
 QY 684 AGGAAAGCGGCTTCTGGGGGGGTGGTGGGGGTCTGAGGAGCTTCAAGTCAAGCGGCTGCTC 743
 DB 601 AGGAAAGCGGCTTCTGGGGGGGTGGTGGGGGTCTGAGGAGCTTCAAGTCAAGCGGCTGCTC 660
 QY 744 AAACAGATCTTCTGCTGCAATTGAACTTCAAGAGCGGCGGCGGCGGCGGCGGCGG 803
 DB 661 AAACAGATCTTCTGCTGCAATTGAACTTCAAGAGCGGCGGCGGCGGCGGCGGCGG 720
 QY 804 AAGGAAAGAGATGAGAGAGCTGAAGTCAAGAGAGAGACGCTCTTGTCTCGATTGAA 863
 DB 721 AAGGAAAGAGATGAGAGAGCTGAAGTCAAGAGAGAGACGCTCTTGTCTCGATTGAA 780

QY 864 CGTATGAAAAGGCGGATGAGCTGTAAAGAGATTAACGAAAGAAAGCAAGCTG 923
 DB 781 CGTATGAAAAGGCGGATGAGCTGTAAAGAGATTAACGAAAGAAAGCAAGCTG 840
 QY 924 TTTCAGGCGTATGAATCTGAAGAGAGAGAGAAACAGAGTATCTGAAGAAATTAATG 983
 DB 841 TTTCAGGCGTATGAATCTGAAGAGAGAGAGAAACAGAGTATCTGAAGAAATTAATG 900
 QY 984 GATGTCACGCGGAGCTTTCCGAGACATCCACATCTGCTCCAAAGCTTCTCATGT 1043
 DB 901 GATGTCACGCGGAGCTTTCCGAGACATCCACATCTGCTCCAAAGCTTCTCATGT 960
 QY 1044 GGGCGAGTGAAGAGGACATTAAGAGAAATCCCAATTTGAAGTACAGAAAGAAAGACT 1103
 DB 961 GGGCGAGTGAAGAGGACATTAAGAGAAATCCCAATTTGAAGTACAGAAAGAAAGACT 1020
 QY 1104 CTTGTTAAAAAGCTGCTCTGAATTTTCAAAAGTCAAAACAAACTCTTAAGACTCT 1163
 DB 1021 CTTGTTAAAAAGCTGCTCTGAATTTTCAAAAGTCAAAACAAACTCTTAAGACTCT 1080
 QY 1164 CTTTAAAGAGAACCTTGTGTTCTTATCTGAACCTTTGTAAAGCTGAATTGAG 1223
 DB 1081 CTTTAAAGAGAACCTTGTGTTCTTATCTGAACCTTTGTAAAGCTGAATTGAG 1140
 QY 1224 AGCCAAAGAAACCCAGAAAGCCCGGCTTCAAGTGAACCCCAAGACTCTCCACT 1283
 DB 1141 AGCCAAAGAAACCCAGAAAGCCCGGCTTCAAGTGAACCCCAAGACTCTCCACT 1200
 QY 1284 CCCCAAAAGGAGACCCAGACCCATCCCAAGAGAAAGCTTCTCAAGTGAATGAAGAT 1343
 DB 1201 CCCCAAAAGGAGACCCAGACCCATCCCAAGAGAAAGCTTCTCAAGTGAATGAAGAT 1260
 QY 1344 TTGCGGTACCTTTTCCACACAGAAATGATTTGTGTGTTGGACACAGCTTCCCATCA 1403
 DB 1261 TTGCGGTACCTTTTCCACACAGAAATGATTTGTGTGTTGGACACAGCTTCCCATCA 1320
 QY 1404 CCGTTACCATTAAGGGAATCTCTCCAAAGAGAGAGAGAGCTGTAGCAAG 1453
 DB 1321 CCGTTACCATTAAGGGAATCTCTCCAAAGAGAGAGAGAGCTGTAGCAAG 1370

RESULT 6
 AA224879
 ID AA224879 standard; DNA; 1509 BP.
 AC
 AC AA224879;
 XX
 DT 02-DEC-1999 (first entry)
 XX
 DE Human secreted protein gene 69 clone HCNK38.
 XX
 XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammatory; ischemic shock; Alzheimer's disease; resection; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 OS Homo sapiens.
 PN W09947540-A1.
 XX
 XX 23-SEP-1999.
 PD
 XX
 PF 18-MAR-1999; 99WC-US005804.
 XX
 PR 19-MAR-1998; 98US-0078563P.
 PR 19-MAR-1998; 98US-0078566P.
 PR 19-MAR-1998; 98US-0078573P.
 PR 19-MAR-1998; 98US-0078574P.

Db 1508 AA 1509

RESULT 7
ADA39700
ADA39700 standard; cDNA; 1509 BP.
ADA39700;
20-NOV-2003 (first entry)
Human secreted protein encoding cDNA.
Human: secreted protein; cancer; hyperproliferative disorder;
rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
anaemia; allergic reaction; asthma; cardiovascular disorder;
wound healing; cytostatic; immunosuppressive; nocitropic; neuroprotective;
antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
vulnerable; cardiac; gene therapy; ss.
Homo sapiens.
MO2002102993-A2.
27-DEC-2002.
19-MAR-2002; 2002MO-US008123.
21-MAR-2001; 2001US-0277340P.
19-JUL-2001; 2001US-0306171P.
13-NOV-2001; 2001US-0331287P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Ruben SM;
WPI; 2003-175238/17.
New human secreted proteins and nucleic acid molecules, useful for
preparing a diagnostic or pharmaceutical composition for diagnosing,
preventing or treating cancer or other hyperproliferative disorder,
asthma, allergies or AIDS.
Claim 9; SEQ ID NO 82; 3205bp; English.
The invention relates to novel genes ADA39629-ADA40565 and proteins
ADA40566-ADA41501 for human secreted proteins, useful for preventing,
treating or ameliorating medical conditions e.g. by protein or gene
therapy. The polypeptides, nucleic acid molecules, antibodies or their
fragments, and agonists or antagonists that bind to the polypeptide are
useful for preparing a diagnostic or pharmaceutical composition for
diagnosing or treating cancer or other hyperproliferative disorder. The
polypeptides and nucleic acid molecules are also useful for detecting,
preventing, diagnosing, prognosticating, treating or ameliorating cancer
or other hyperproliferative disorders including neoplasms, autoimmune
disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
anaemia), haematopoietic or haematological disorders (e.g. anaemia,
thrombocytopenia), allergic reactions including asthma or eczema,
inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
Alzheimer's disease or Parkinson's disease), cardiovascular disorders
(e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
fungal or viral infections including HIV/AIDS), or wound healing and
disorders of epithelial cell proliferation. The nucleic acids are also
useful for chromosome identification, radiation hybrid mapping or long-
range restriction mapping, as molecular weight markers, or as
hybridization or diagnostic probes. The polypeptides and antibodies are
useful for providing immunohistochemical probes for differential identification
of the tissues immunohistochemistry assays. Note: The sequence data for
this patent did not form part of the printed specification, but was
obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1509 BP; 397 A; 272 C; 319 G; 521 T; 0 U; 0 Other;
SQ
Query Match 28.8%; Score 1260; DB 8; Length 1509;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1500; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 2859 GATGACCTAATGAGCTTCTCCATTCACCTTGTAAAAATATTTGTATGTATACCATCTT
DB 9 GATGACCTAATGAGCTTCTCCATTCACCTTGTAAAAATATTTGTATGTATACCATCTT
QY 2919 GGTCTCTCCCTCCCGTTTGTAAAAATATTCAGATGACCTCCAGCCACTTGTGTC
DB 69 GGTCTCTCCCTCCCGTTTGTAAAAATATTCAGATGACCTCCAGCCACTTGTGTC
QY 2979 TCAGTGTAAGATCCCTTATTAATATCTGAAAGAAAAATAGCCAGACCTCTGTCTCA
DB 129 TCAGTGTAAGATCCCTTATTAATATCTGAAAGAAAAATAGCCAGACCTCTGTCTCA
QY 3039 AATATATAGGAATGCTTCTTATGCTCAGAGACTATTTGTGAAACAAAGTAGGGGT
DB 189 AATATATAGGAATGCTTCTTATGCTCAGAGACTATTTGTGAAACAAAGTAGGGGT
QY 3099 CTATCTCTAGAGAGGTAGGGGCTTTATCTTAAAGAAATATGTCGCCAGATTATTA
DB 249 CTATCTCTAGAGAGGTAGGGGCTTTATCTTAAAGAAATATGTCGCCAGATTATTA
QY 3159 CACTTTTAAAGAGAGAGCCAAAGGTATGTA-GGGTGTGTGCTGGCCCATCATGTGAGCAC
DB 308 CACTTTTAAAGAGAGAGCCAAAGGTATGTA-GGGTGTGTGCTGGCCCATCATGTGAGCAC
QY 3218 GAAGAGAGATGGGATTCATTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
DB 368 GAAGAGAGATGGGATTCATTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
QY 3278 GCTAAAGTTTGTGAGATGTTGATCTGTCTCTGATTTGACTTTTAAAGAAATTA
DB 428 GCTAAAGTTTGTGAGATGTTGATCTGTCTCTGATTTGACTTTTAAAGAAATTA
QY 3338 TTCTGGCAGACATGTATGTAATCTTGATGATCTGTCTCTTATTTCTTTGTGT
DB 488 TTCTGGCAGACATGTATGTAATCTTGATGATCTGTCTCTTATTTCTTTGTGT
QY 3398 GT
DB 548 GT
QY 3458 GGGCTCTCTTAAAGAGAGAGCTGTGTAATCTTATTCAGCAAGATGTAGAGAAATA
DB 608 GGGCTCTCTTAAAGAGAGAGCTGTGTAATCTTATTCAGCAAGATGTAGAGAAATA
QY 3518 GGAATTATTCACCTAGAGGCTCTCATCTCAACCTTAAAGAGAGATTTCTAGAAAAAC
DB 668 GGAATTATTCACCTAGAGGCTCTCATCTCAACCTTAAAGAGAGATTTCTAGAAAAAC
QY 3578 TGGGCCAGATTTTCTTTGTTCTTCATATTTTATGTGCGAGGCTGTTCAGTTTCTTAC
DB 728 TGGGCCAGATTTTCTTTGTTCTTCATATTTTATGTGCGAGGCTGTTCAGTTTCTTAC
QY 3638 TCTTACCTATGTATATTTCTTGTGTAAGTGTCCAAAAAGAAAAAGACCAATCAGTGT
DB 788 TCTTACCTATGTATATTTCTTGTGTAAGTGTCCAAAAAGAAAAAGACCAATCAGTGT
QY 3698 CTCTTACCTTGTCTTGTGTAATCCCTAGTTCTTCTTGTATTTGAGATGTGTGCTTCC
DB 848 CTCTTACCTTGTCTTGTGTAATCCCTAGTTCTTCTTGTATTTGAGATGTGTGCTTCC
QY 3758 TAATTTTGGTATGATGATTAACAAATTTAAACCAATTTGTTGTGCTTAAACCAAGGAGCTC
DB 908 TAATTTTGGTATGATGATTAACAAATTTAAACCAATTTGTTGTGCTTAAACCAAGGAGCTC
QY 3818 CCGAGTTTCTGACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT

[illegible][illegible]

[illegible]

Db	1388	TGTTGTAATTCATACACTTTTGATACACATTCGANGGTGTAATAATTGGTGTGCTGTAAT	1447
Qy	4298	ATCTATAAAGAGGTCGAATTTGTAATAAATACTATGTCGCTGTAAAAA	4357
Db	1448	ATCTATATAAAGAGTTCATATTGTAATAATACTATGTCGCTGTAAAAA	1507
Qy	4358	AA 4359	
Db	1508	AA 1509	
RESULT 10			
ADAS5894	ID	ADAS5894 standard; DNA; 1509 BP.	
XX	AC	ADAS5894;	
XX	DT	20-NOV-2003 (first entry)	
XX	DE	Gene encoding human secreted protein #73.	
XX	XX	immunosuppressive; antiinflammatory; antiasthmatic; anti allergic;	
KW	XX	cytostatic; cerebroprotective; neuroprotective; neurotropic;	
KW	KW	cardiovascular; antiarteriosclerotic; gene therapy;	
KW	KW	human secreted protein; immune disorder; inflammation;	
KW	KW	respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;	
KW	KW	inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;	
KW	KW	multiple sclerosis; ischemic brain injury; Parkinson's disease;	
KW	KW	Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;	
KW	KW	triple helix formation; antisense gene therapy; forensic biology; ds;	
KM	gene.		
XX	XX		
OS	XX	Homo sapiens.	
XX	XX		
PN	XX	WC0002102994-A2.	
PD	XX	27-DEC-2002.	
PF	XX	19-MAR-2002; 2002MO-US008278.	
PR	XX	21-MAR-2001; 2001US-0277340P.	
PR	XX	19-JUL-2001; 2001US-0306171P.	
PR	XX	13-NOV-2001; 2001US-0331287P.	
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	XX	Rosen CA, Ruben SM;	
DR	XX	WPI; 2003-167512/16.	
DR	XX	P-PSDB; ADAS6791.	
PT	XX	New human secreted polypeptides and polynucleotides, useful for	
PT	XX	diagnosing, treating or preventing e.g. immune disorders, inflammatory	
PT	XX	conditions, respiratory disorders, cancers, CNS disorders, or	
PT	XX	neurodegenerative disorders.	
PS	XX	Claim 21; SEQ ID NO 83; 1754dp; English.	
CC	XX	The invention relates to 592 new human secreted polypeptides useful for	
CC	XX	diagnosing, treating or preventing e.g. immune disorders, inflammatory	
CC	XX	conditions, respiratory disorders, cancers, CNS disorders, or	
CC	XX	neurodegenerative disorders, or polypeptides comprising an amino acid	
CC	XX	sequence at least 95% identical to the new sequences. The polypeptides,	
CC	XX	antibodies or antibody fragments that bind to the polypeptides, nucleic	
CC	XX	acids encoding the polypeptides, agonists or antagonists that binds to	
CC	XX	the polypeptide, are useful in preparing diagnostic or pharmaceutical	
CC	XX	compositions for diagnosing, treating or preventing an e.g. immune	
CC	XX	disorders, inflammatory conditions (e.g. inflammatory bowel disease,	
CC	XX	nephritis or Crohn's disease), respiratory disorders (e.g. asthma and	
CC	XX	allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders	
CC	XX	(e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative	
CC	XX	disorders (e.g. Parkinson's disease or Alzheimer's disease), and	
CC	XX	cardiovascular disorders (e.g. atherosclerosis or myocarditis). The	

polynucleotides are useful for chromosome identification, chromosome mapping, for controlling gene expression through triple helix formation or antisense DNA or RNA, in gene therapy, for identifying individuals from minute biological samples, in forensic biology, and as hybridization probes. The polypeptides are useful for as molecular weight markers on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels, to raise antibodies, for testing biological activities, and for treating or preventing neural disorders, immune system disorders, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, proliferative and/or cancerous diseases. This sequence corresponds to a gene encoding one of the polypeptide of the invention. Note: The sequence data for this patent did form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pct_sequences.

SQ Sequence 1509 BP; 397 A; 272 C; 319 G; 521 T; 0 U; 0 Other;

Query Match	28.8%;	Score 1260;	DB 10;	Length 1509;
-------------	--------	-------------	--------	--------------

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1500; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY	2859	GGGTGACTCTAATGAGCTTCCTCCATCCACTCTTGTGTAATAATATTGTGATNGGTACCATCTT	2918
Db	9	GAGGTACCTTAATGAGCTTCCTCCATCCACTCTTGTGTAATAATATTGTGATNGGTACCATCTT	68
QY	2919	GGTCTCTCCCTCCCGTTTGTGTAATAATACAGAAATGACACTCCAGGGCACCTTGGCT	2978
Db	69	GGTCTCTCCCTCCCGTTTGTGTAATAATACAGAAATGACACTCCAGGGCACCTTGGCT	128
QY	2979	TCAGTGTAAAGATCCCTATTAACTATCTGAAAGAAAATAGAGCCAGACCTTGTCTCA	3038
Db	129	TCAGTGTAAAGATCCCTATTAACTATCTGAAAGAAAATAGAGCCAGACCTTGTCTCA	188
QY	3039	AATATATAAGAAATTCCTTCTTTAGTCTTAGAGCTATTTGTGTGAAACAGTAGGGGT	3098
Db	189	AATATATAAGAAATTCCTTCTTTAGTCTTAGAGCTATTTGTGTGAAACAGTAGGGGT	248
QY	3099	CTAATCTCCCTAAGAGTAGGGGCTTTTATCTTTAAAGAAATATGTCCCGCAGATTATAG	3158
Db	249	CTAATCTCCCTAAGAGTAGGGGCTTTTATCTTTAAAGAAATATGTCCCGCAGATTATAG	307
QY	3159	CACCTTTAGAGAGAAAGCCAAAGGATATGTA-GGGTGTGTGGCTGGCCATCAGTGGAGCAC	3217
Db	308	CACCTTTAGAGAGAAAGCCAAAGGATATGTAAGGGGTGTGTGGCTGGCCATCAGTGGAGCAC	367
QY	3218	GAAAGAGAAATGGGATACCATTTGTGGGAGAAAGAAAGAAAGTTTCTCAGGGGCTCCCACT	3277
Db	368	GAAAGAGAAAGGAGATACCATTTGTGGGAGAAAGAAAGAAAGTTTCTCAGGGGCTCCCACT	427
QY	3278	GCTAAAGTTTTTGTGAGATGTGATCTGATCTGACCTCCGAGTTTGACTTTTAAAGGAATTA	3337
Db	428	GCTAAAGTTTTTGTGAGATGTGATCTGATCTGACCTCCGAGTTTGACTTTTAAAGGAATTA	487
QY	3338	TTCTGGAGCAGCATGTAGATATTCTTGTGATGATCTGTGCTCTTATTTCTCCTTTTGTGT	3397
Db	488	TTCTGGAGCAGCATGTAGATATTCTTGTGATGATCTGTGCTCTTATTTCTCCTTTTGTGT	547
QY	3398	GTGTGTGTGTGTGTGTGTGCTATGAGGTTTTCAATTGTAACTCCATCTGCTTTAGAGAGT	3457
Db	548	GTGTGTGTGTGTGTGTGTGCTATGAGGTTTTCAATTGTAACTCCATCTGCTTTAGAGAGT	607
QY	3458	GGGCTCTCTAATAAGGAACCTGCTGTAAACTTCATTCGAGAGAGAGATGTAAAGGAATA	3517
Db	608	GGGCTCTCTAATAAGGAACCTGCTGTAAACTTCATTCGAGAGAGAGATGTAAAGGAATA	667
QY	3518	GGACTTAAATTCACCTAAGGGGCTCTATCTCAACACCTTAAAGAGAGAGATTTCTAGAAAAAC	3577
Db	668	GGACTTAAATTCACCTAAGGGGCTCTATCTCAACACCTTAAAGAGAGAGATTTCTAGAAAAAC	727
QY	3578	TGGGCGAGATTTTCTTTGTGTCCATCATATTTTAATGTGCGAGGCTGTCACTTTTCTTAC	3637
Db	728	TGGGCGAGATTTTCTTTGTGTCCATCATATTTTAATGTGCGAGGCTGTCACTTTTCTTAC	787

QY	3638	TCCTACCTATGTAATATTTCTTGTAACGTGTCCAAAAAGAAAAAGAACCAATCACTGT	3697
Db	788	TCCTACCTATGTAATATTTCTTGTAACGTGTCCAAAAAGAAAAAGAACCAATCACTGT	847
QY	3698	CTCTTGACTTTGTTCTTTGATCCCTCAGTTTCTTTGATTTGACAGTGTGTGGGTTCC	3757
Db	848	CTCTTGACTTTGTTCTTTGATCCCTCAGTTTCTTTGATTTGACAGTGTGTGGGTTCC	907
QY	3758	TAAATTTGGGTAAGAGTTAGCAAAATTTAACCAATTTGTGTGTGCCCTTACCCAGGGAGTC	3817
Db	908	TAAATTTGGGTAAGAGTTAGCAAAATTTAACCAATTTGTGTGTGCCCTTACCCAGGGAGTC	967
QY	3818	CCCAAGTTTCGACCTGTAAGTGAACCTGGAAGAAATCCAGAGGTCTATCGGCCAGATT	3877
Db	968	CCCAAGTTTCGACCTGTAAGTGAAGTGAAGAAATCCAGAGGTCTATCGGCCAGATT	1027
QY	3878	AAGTAGATTTATTTCTTGTTGTCTCCCTCCCTGAGAGCTCTTATTTTATTTATGTCCC	3937
Db	1028	AAGTAGATTTATTTCTTGTTGTCTCCCTCCCTGAGAGCTCTTATTTTATTTATGTCCC	1087
QY	3938	TCCTTAGAGTTAATTCCTCTTTGATTTGACTTTGTTGAGAAGAGGTTTGAACAGTAGATT	3997
Db	1088	TCCTTAGAGTTAATTCCTCTTTGATTTGACTTTGTTGAGAAGAGGTTTGAACAGTAGATT	1147
QY	3998	AGCAAAATTCAGTGAAGAAATTTACAGTGTGTAGAGTGTGGGGGAAATTTAGCTTAT	4057
Db	1148	AGCAAAATTCAGTGAAGAAATTTACAGTGTGTAGAGTGTGGGGGAAATTTAGCTTAT	1207
QY	4058	TTTTCCCTACATGGGATACAAACACTGTGAATTCATCTTAACTGAAGGCCCTGCAGTTC	4117
Db	1208	TTTTCCCTACATGGGATACAAACACTGTGAATTCATCTTAACTGAAGGCCCTGCAGTTC	1267
QY	4118	TCCTAAAAACATAGTTGTTGTTTCTTTTAAACAAGTTTAAAGTTAATTAATTA	4177
Db	1268	TCCTAAAAACATAGTTGTTGTTTCTTTTAAACAAGTTTAAAGTTAATTAATTA	1327
QY	4178	AAAAAATTCCTGTCTGTCTACTTACCTCAGCTTGTGTTTATGCCCATTTCAATGTGTCTG	4237
Db	1328	AAAAAATTCCTGTCTGTCTACTTACCTCAGCTTGTGTTTATGCCCATTTCAATGTGTCTG	1387
QY	4238	TGTTGTAAATCATPACTTTTGATTCACATTTCTGATGTGTAAATTTGGTGTCTGTAAAT	4297
Db	1388	TGTTGTAAATCATPACTTTTGATTCACATTTCTGATGTGTAAATTTGGTGTCTGTAAAT	1447
QY	4298	ATCTTATTAAGAGGTTCAATTTGTAATTAACATATTTGGCTGTAAAAA	4357
Db	1448	ATCTTATTAAGAGGTTCAATTTGTAATTAACATATTTGGCTGTAAAAA	1507
QY	4358	AA 4359	
Db	1508	AA 1509	
RESULT 11			
ADI30841			
ID	ADI30841	standard; cDNA; 1545 BP.	
XX	ADI30841;		
AC	17-JUN-2004	(first entry)	
DT	Human cDNA #167.		
DE			
XX			
KW	Human; gene; ss; immunological response; immunopathological condition;		
KW	Crohn's disease; aetna; ulcerative colitis; hypersensitophilia;		
KW	irritable bowel syndrome; osteoarthritis; rheumatoid arthritis;		
KW	acute monocytic leukaemia; antiinflammatory; antiasthmatic; antiulcer;		
KW	osteopathic; antiarthritic; antirheumatic; cyostatic.		
XX			
OS	Homo sapiens.		
XX			
FN	US6607879-B1		
XX			

PD 19-AUG-2003.
XX
XX 09-FEB-1998; 98US-00023655.
XX
XX 09-FEB-1998; 98US-00023655.
XX
XX (INCY-) INCYTE CORP.
XX
XX Cocke BG, Stuart SG, Seilhamer JJ;
XX WPI; 2003-895307/82.
XX
XX A composition comprising a plurality of cDNAs, useful for detecting
PT altered expression of genes in an immunological response or for
PT diagnosing and treating an immunopathology, e.g. Crohn's disease, asthma
PT or osteoarthritis.
PS
PS Claim 1; SEQ ID NO 167; 50bp; English.

CC The invention relates to a composition comprising a plurality of cDNAs
CC for detecting the altered expression of genes in an immunological
CC response. The invention also relates to a method of diagnosing or
CC monitoring the treatment of an immunopathological condition in a sample,
CC comprising obtaining nucleic acids from a sample, contacting the nucleic
CC acids of the sample with an array comprising the plurality of cDNAs under
CC conditions to form one or more hybridization complexes, detecting the
CC hybridization complexes and comparing the levels of the detected
CC hybridization complexes with the level of hybridization complexes
CC detected in a non-diseased sample, where an altered level of the detected
CC immunopathological condition. Also disclosed are an expression profile
CC comprising a microarray and a plurality of detectable complexes and a
CC method for identifying a plurality of polynucleotide probes. The cDNAs
CC are useful as hybridizable array elements in a microarray for monitoring
CC the expression of target polynucleotides. The microarray can be used in
CC the diagnosis of an immunopathology, such as Crohn's disease, asthma,
CC ulcerative colitis, hypersplenophilia, irritable bowel syndrome, osteo-
CC osteoarthritis, rheumatoid arthritis or acute monocytic leukemia, and in
CC identifying agents for the treatment of the diseases. The microarray may
CC also be used in drug discovery and development, toxicological and
CC carcinogenicity studies, forensics or pharmacogenomics. The composition
CC may also be used in purification of a subpopulation of mRNAs, cDNAs or
CC genomic fragments. This sequence represents a human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX

Sequence 1545 BP; 392 A; 279 C; 318 G; 521 T; 0 U; 35 Other;

Query Match 20.7%; Score 903; DB 11; Length 1545;
Best Local Similarity 100.0%; Pred. No. 5.5e-313;
Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3426 TTTCATTGTTGAATCCATCTGCTTAGAGAGTGGGCTCTGTATTAAGGAACTGCTGTAA 3485
DB 631 TTTCATTGTTGAATCCATCTGCTTAGAGAGTGGGCTCTGTATTAAGGAACTGCTGTAA 690
QY 3486 ACTTCATTGAGAGAGATGTAGAGAAATAGGACTTAATTCACATGAGGCTCTCATC 3545
DB 691 ACTTCATTGAGAGAGATGTAGAGAAATAGGACTTAATTCACATGAGGCTCTCATC 750
QY 3546 TCACACCTTAAGAGAGATTTCTAGAAAACTGGGCCAGATTTTCTTTGTTCTCATCA 3605
DB 751 TCACACCTTAAGAGAGATTTCTAGAAAACTGGGCCAGATTTTCTTTGTTCTCATCA 810
QY 3606 TTTTAATGTCGAGGCTGTCAGTTTCTTACTCTTACCTTAACGATGATTTCTTCGTAAC 3665
DB 811 TTTTAATGTCGAGGCTGTCAGTTTCTTACTCTTACCTTAACGATGATTTCTTCGTAAC 870
QY 3666 GTGTCCAAAAAGAAAAAGACCAATCAGTGTCTCTGACTTTGTTCTTTGATCCCTCAG 3725
DB 871 GTGTCCAAAAAGAAAAAGACCAATCAGTGTCTCTGACTTTGTTCTTTGATCCCTCAG 930

QY 3726 TTTCTTCTTGATTTTGAGAGATGTGGGTTCCCTAATTTTGGGATAGTAGAAATTTA 3785
DB 931 TTTCTTCTTGATTTTGAGAGATGTGGGTTCCCTAATTTTGGGATAGTAGAAATTTA 990
QY 3786 ACCATTGTGTGTGGCCCTACCCAGGGGACTCCCGATTTCTGACTTGAAGTACTGAG 3845
DB 991 ACCATTGTGTGTGGCCCTACCCAGGGGACTCCCGATTTCTGACTTGAAGTACTGAG 1050
QY 3846 AAGATTCACAGAGTGTCTATCTGGCCAGATTTAAGTAGATTTCTTGGTTCTCCC 3905
DB 1051 AAGATTCACAGAGTGTCTATCTGGCCAGATTTAAGTAGATTTCTTGGTTCTCCC 1110
QY 3906 TCTCCCTGAGGACCTCTTATTTTATTTATGTCCTCTCTGATTTATTTCTTGGATTTG 3965
DB 1111 TCTCCCTGAGGACCTCTTATTTTATTTATGTCCTCTCTGATTTATTTCTTGGATTTG 1170
QY 3966 ACTTTGTTGAGAGAGGTTTGAACAGTAGATTAGCAAGTTCACAGTGCAGAAATTAACAT 4025
DB 1171 ACTTTGTTGAGAGAGGTTTGAACAGTAGATTAGCAAGTTCACAGTGCAGAAATTAACAT 1230
QY 4026 GTGTAGAGTGTGGGGGAAATTAATGATTTATTTTCCCTACATGGGATACACACTGTG 4085
DB 1231 GTGTAGAGTGTGGGGGAAATTAATGATTTATTTTCCCTACATGGGATACACACTGTG 1290
QY 4086 AATTCAATCTTCAACTGAAAGCCCTGACGTTCTCTAAACATAGTTGTTTCTT 4145
DB 1291 AATTCAATCTTCAACTGAAAGCCCTGACGTTCTCTAAACATAGTTGTTTCTT 1350
QY 4146 TAACAAGTTTAAAGTAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4205
DB 1351 TAACAAGTTTAAAGTAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1410
QY 4206 TTTGTTTATGCCCATTTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 4265
DB 1411 TTTGTTTATGCCCATTTTCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1470
QY 4266 TTTGATGTTAAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 4325
DB 1471 TTTGATGTTAAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1530
QY 4326 ACT 4328
DB 1531 ACT 1533

RESULT 12
AAF18193
ID AAF18193 standard; DNA; 1529 BP.

AC AAF18193;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 212.
XX
KW Human; lung cancer associated protein; neuroprotective; cyrostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephroretropic; anticinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
PN MO200055180-A2.
XX
PD 21-MAR-2000.
XX
PF 08-MAR-2000; 2000MO-US005918.
XX
PR 12-MAR-1999; 99US-0124270P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.

XX Ruben SM;
 XX WPI: 2000-587514/55.
 DR P-PSDB: AAB58317.
 XX Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 XX such as lung cancer.
 PS Claim 1; Page 676-677; 1425bp; English.
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective, cytostatic, cardioactive;
 CC immunomodulatory; muscular active general; vulnery; gastrointestinal
 CC general; nephrotoxic; anti-infective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the protein
 CC or polynucleotide sequences. The lung cancer associated polynucleotide
 CC sequences may be used for detection of lung cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The proteins may be used to treat disorders such as
 CC neutral, immune, muscular, reproductive, gastrointestinal, pulmonary,
 CC cardiovascular, renal, and proliferative disorders. The proteins may also
 CC be used in the treatment of wounds and infectious diseases.
 CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
 CC used in the course of the invention for the identification and
 CC characterisation of the polynucleotide and protein sequences
 CC
 XX
 SQ Sequence 1529 BP; 460 A; 344 C; 379 G; 344 T; 0 U; 2 Other;
 Query Match 19.1%; Score 837; DB 3; Length 1529;
 Best Local Similarity 100.0%; Pred. No. 2.2e-289; Indels 0; Gaps 0;
 Matches 837; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 617 GCCACCCCTCGGCCCAACCCGCGCGGAGCCCTGCGCGCGACGAGGCGAGATGGA 676
 DB 2 GCCACCCCTCGGCCCAACCCGCGCGGAGCCCTGCGCGCGACGAGGCGAGATGGA 61
 QY 677 GAGTATGAGAGAGCCCTCTCGGAGGCTGCGCGCTCGGAGGCTTCAGTCAAGCCGC 736
 DB 62 GAGTATGAGAGAGCCCTCTCGGAGGCTGCGCGCTCGGAGGCTTCAGTCAAGCCGC 121
 QY 737 CTGCTCAAAAGATCCTTCTGCTGCAATTGCACTCAATCAAGAGAGAGAGAGAGAG 796
 DB 122 CTGCTCAAAAGATCCTTCTGCTGCAATTGCACTCAATCAAGAGAGAGAGAGAGAG 181
 QY 797 GCAGGCCAAGAGAGAGATCGAGAGCTGAAGTCAGAGAGAGAGAGAGAGAGAGAGAG 856
 DB 182 GCAGGCCAAGAGAGAGATCGAGAGCTGAAGTCAGAGAGAGAGAGAGAGAGAGAGAG 241
 QY 857 GATTGAACGTATGAGAGAGCGGATGCAAGCTGTAAAGAGATTAAGAGAGAGAGAG 916
 DB 242 GATTGAACGTATGAGAGAGCGGATGCAAGCTGTAAAGAGATTAAGAGAGAGAGAG 301
 QY 917 CAAGCTGTTTCAAGGCTATGAACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 976
 DB 302 CAAGCTGTTTCAAGGCTATGAACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
 QY 977 TAAACTGAGAGCGAGCGAGCTTTCCGAGAGATCCAGAGCTGCTCCCAAGCCCTT 1036
 DB 362 TAAACTGAGAGCGAGCGAGCTTTCCGAGAGATCCAGAGCTGCTCCCAAGCCCTT 421
 QY 1037 CTCATGCGCGAG 1096
 DB 422 CTCATGCGCGAG 481
 QY 1097 AAAAGCTCCCTTAAAG 1156
 DB 482 AAAAGCTCCCTTAAAG 541
 QY 1157 GCACTCTCTATTAAG 1216

DB 542 GCACTCTCTATTAAG 601
 QY 1217 ATTGAGAGCGAG 1276
 DB 602 ATTGAGAGCGAG 661
 QY 1277 CTCACCTCCCAAAAG 1336
 DB 662 CTCACCTCCCAAAAG 721
 QY 1337 AGAAGATTGCGCGTACCTTTCCACACAGAAATATTTGTGTGTGGACCGCTCC 1396
 DB 722 AGAAGATTGCGCGTACCTTTCCACACAGAAATATTTGTGTGTGGACCGCTCC 781
 QY 1397 CCCATCACCGTTACATTAGGGGAATCTCTCCAAAGAGAGAGAGAGAGAGAGAG 1453
 DB 782 CCCATCACCGTTACATTAGGGGAATCTCTCCAAAGAGAGAGAGAGAGAGAGAG 838
 RESULT 13
 AAC76874
 ID AAC76874 standard; cDNA; 2887 BP.
 XX
 AC AAC76874;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2429 polynucleotide sequence SEQ ID NO:4857.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipapillary; antipapillary; antipapillary; antipapillary;
 KW anticonvulsant; osteoporotic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; anti-inflammatory;
 KW antiviral; antibacterial; antifungal; antihypertensive; antihypertensive;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; anti-inflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 KW
 XX
 OS Homo sapiens.
 XX
 PN MO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US008621.
 XX
 PR 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB: AAB42665.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 5; Page 404-4045; 5507bp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipariatic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
CC anticonvulsant; antihistytic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antineoplastic; antibacterial;
CC antiviral; antifungal; antineumatic; antihydroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORF-associated disorder. The nucleic acids can be used to express ORF
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, anti-inflammatory disease; to
CC enhance coagulation; to inhibit thrombosis, and as a contraceptive
XX
SQ Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 0 U; 2 Other;

Query Match 16.7%; Score 732; DB 3; Length 2887;
Best Local Similarity 100.0%; Pred. No. 6.6e-252;
Matches 732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 722 CTCCTAGTCAGGCGGCTGCTCAACAGATCTTCTGCAATTGGACCTCATCGACA 781
DB 69 CTCCTAGTCAGGCGGCTGCTCAACAGATCTTCTGCAATTGGACCTCATCGACA 128
QY 782 GCAGCAGCAGCAGCTGCGAGCCAAAGAAAAGAGATGAGAGCTGAAGTCAGAGAGAGA 841
DB 129 GCAGCAGCAGCAGCTGCGAGCCAAAGAAAAGAGATGAGAGCTGAAGTCAGAGAGAGA 188
QY 842 CACGCTCTTGTCTCGAGTTGACGTATGAAAAGCGGATGCACTGGTAAAGAGATAA 901
DB 189 CACGCTCTTGTCTCGAGTTGACGTATGAAAAGCGGATGCACTGGTAAAGAGATAA 248
QY 902 CGAGAAAAGAAAAGCACAAGCTGTTTCAGGGCTATGAAACTGAGAGAGAGAGAAACAGA 961
DB 249 CGAGAAAAGAAAAGCACAAGCTGTTTCAGGGCTATGAAACTGAGAGAGAGAGAAACAGA 308
QY 962 GCTATCTGAGAAAATTAACTGAGAGTGCAGCGGAGCTTCCGAGACATCCGAGACTCT 1021
DB 309 GCTATCTGAGAAAATTAACTGAGAGTGCAGCGGAGCTTCCGAGACATCCGAGACTCT 368
QY 1022 GCCTCCCAAGCCCTTCTCATGTGGCGGAGTGGAAAAGGACATTAAGAATCCCATTT 1081
DB 369 GCCTCCCAAGCCCTTCTCATGTGGCGGAGTGGAAAAGGACATTAAGAATCCCATTT 428
QY 1082 TGGAAAGTACGAAAAGAAAAGCTCTGTTAAAAAGCTGCTCTGAATTTTCAAAAAGTCAA 1141
DB 429 TGGAAAGTACGAAAAGAAAAGCTCTGTTAAAAAGCTGCTCTGAATTTTCAAAAAGTCAA 488
QY 1142 AACAAAACCTCTAAGCAGCTCTCTATTAAAGGAAACCTGTGGTCCCTTATGTGAAC 1201
DB 489 AACAAAACCTCTAAGCAGCTCTCTATTAAAGGAAACCTGTGGTCCCTTATGTGAAC 548
QY 1202 TGTTTTGAACGTGAATTGAGAGCCAAAGAAACCCCAAGAAAGCCCGCTTCAAGTGA 1261
DB 549 TGTTTTGAACGTGAATTGAGAGCCAAAGAAACCCCAAGAAAGCCCGCTTCAAGTGA 608
QY 1262 CACCCCAACCAAGACTCTCACTCCCAAAAAGGACCCAGACCCCATCCCAAGAGAAAGC 1321
DB 609 CACCCCAACCAAGACTCTCACTCCCAAAAAGGACCCAGACCCCATCCCAAGAGAAAGC 668
QY 1322 CTTCTCAAGTGAATGAGATTTGCGGTACCTTTTCAACACAGAAATGATTTGTGTCG 1381
DB 669 CTTCTCAAGTGAATGAGATTTGCGGTACCTTTTCAACACAGAAATGATTTGTGTCG 728
QY 1382 TTGGCACAACAGCTCCCATCACCGTTACATTTACGGGAAATCTCTCAAGAAAGAGAGA 1441
DB 729 TTGGCACAACAGCTCCCATCACCGTTACATTTACGGGAAATCTCTCAAGAAAGAGAGA 788

QY 1442 GACTGTAGCAG 1453
DB 789 GACTGTAGCAG 800

RESULT 14

ID ADI30996 standard; cDNA; 1290 BP.

ADI30996;

17-JUN-2004 (first entry)

Human cDNA #322.

XX Human; gene; ss; immunological response; immunopathological condition;
XX Crohn's disease; asthma; ulcerative colitis; hypersensitivity;
XX irritable bowel syndrome; osteoarthritis; rheumatoid arthritis;
XX acute monocytic leukemia; antineoplastic; antineoplastic; antineoplastic;
XX osteopathic; antineoplastic; antineoplastic; antineoplastic.

XX Homo sapiens.

XX US6607879-B1.

XX 19-ANG-2003.

XX 09-FEB-1998; 98US-00023655.

XX 09-FEB-1998; 98US-00023655.

XX (INCY-) INCYTE CORP.

XX Cocks BG, Stuart SG, Selthamer JI;

XX WPI; 2003-895307/82.

PT A composition comprising a plurality of cDNAs, useful for detecting
PT altered expression of genes in an immunological response or for
PT diagnosing and treating an immunopathology, e.g. Crohn's disease, asthma
PT or osteoarthritis.

XX Claim 1; SEQ ID NO 322; 50pp; English.

CC The invention relates to a composition comprising a plurality of cDNAs
CC for detecting the altered expression of genes in an immunological
CC response. The invention also relates to a method of diagnosing or
CC monitoring the treatment of an immunopathological condition in a sample,
CC comprising obtaining nucleic acids from a sample, contacting the nucleic
CC acids of the sample with an array comprising the plurality of cDNAs under
CC conditions to form one or more hybridisation complexes, detecting the
CC hybridisation complexes and comparing the levels of the detected
CC hybridisation complexes with the level of hybridisation complexes
CC detected in a non-diseased sample, where an altered level of the detected
CC hybridisation complexes correlates with the presence of an
CC immunopathological condition. Also disclosed are an expression profile
CC comprising a microarray and a plurality of detectable complexes and a
CC method for identifying a plurality of polynucleotide probes. The cDNAs
CC are useful as hybridisable array elements in a microarray for monitoring
CC the expression of target polynucleotides. The microarray can be used in
CC the diagnosis of an immunopathology, such as Crohn's disease, asthma,
CC ulcerative colitis, hypersensitivity, irritable bowel syndrome,
CC osteoarthritis, rheumatoid arthritis or acute monocytic leukaemia, and in
CC identifying agents for the treatment of the diseases. The microarray may
CC also be used in drug discovery and development, toxicological and
CC carcinogenicity studies, forensics or pharmacogenomics. The composition
CC may also be used in purification of a subpopulation of mRNAs, cDNAs or
CC genomic fragments. This sequence represents a human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html.

SQ Sequence 1290 BP; 390 A; 275 C; 304 G; 321 T; 0 U; 0 Other;

Query Match 11.1%; Score 487; DB 11; Length 1290;
Best Local Similarity 99.8%; Pred. No. 2.8e-164;
Matches 607; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 847 TCCTGCTCGGATTGAACGATGAGAAAGCGGATGACGCTGGTAAAGAGATACAGAGA 906
DB 1 TCCTGCTCGGATTGAACGATGAGAAAGCGGATGACGCTGGTAAAGAGATACAGAGA 60
QY 907 AAGAAAGGCAAGCTGTTTCAGGGCTATGAAATGAAAGAGAGAGAAACAGAGCTAT 966
DB 61 AAGAAAGGCAAGCTGTTTCAGGGCTATGAAATGAAAGAGAGAGAAACAGAGCTAT 120
QY 967 CTGGAAGAAATTAATCTGAGTGCAGCGGAGCTTTCCGAGACATCCAGACTCTGCTC 1026
DB 121 CTGGAAGAAATTAATCTGAGTGCAGCGGAGCTTTCCGAGACATCCAGACTCTGCTC 180
QY 1027 CCAAGCCCTTCTCATGTGGGCGAGT-GGAAAGGACATAAAGAAATCCCATTTGGA 1085
DB 181 CCAAGCCCTTCTCATGTGGGCGAGTGGAAAGGACATAAAGAAATCCCATTTGGA 240
QY 1086 AGTACAGAAAGAAAGACTCTCTGTTAAAGAGTGGCTCTGAAATTTCAAAAGTCAAAACA 1145
DB 241 AGTACAGAAAGAAAGACTCTCTGTTAAAGAGTGGCTCTGAAATTTCAAAAGTCAAAACA 300
QY 1146 AAAAATCTTAAGCACTCTCTATTAAGAGAACCCCTGGTCTTATCTGAAAGCTT 1205
DB 301 AAAAATCTTAAGCACTCTCTATTAAGAGAACCCCTGGTCTTATCTGAAAGCTT 360
QY 1206 TGTAAAGTGAATGAGAGGCAAGAAACCCAGAGAAAGCCCGCTTCTGATGAGACACC 1265
DB 361 TGTAAAGTGAATGAGAGGCAAGAAACCCAGAGAAAGCCCGCTTCTGATGAGACACC 420
QY 1266 CCACCAAGACTCTCTACCTCCCAAAAGGAGCCGACCCATCCCAAGAGAGAAAGCTTC 1325
DB 421 CCACCAAGACTCTCTACCTCCCAAAAGGAGCCGACCCATCCCAAGAGAGAAAGCTTC 480
QY 1326 TCAAGTGAATGAGAGATTTGGCGTACCTTTCCACACAGAAATGTTTGTGTGG 1385
DB 481 TCAAGTGAATGAGAGATTTGGCGTACCTTTCCACACAGAAATGTTTGTGTGG 540
QY 1386 CACCAAGCTCCCGCATCAGCTTACATTCAGGAAATCTCTCCAAAGAGAGAGACT 1445
DB 541 CACCAAGCTCCCGCATCAGCTTACATTCAGGAAATCTCTCCAAAGAGAGAGACT 600
QY 1446 GTAGCAAG 1453
DB 601 GTAGCAAG 608

RESULT 15
AA193984/c
ID AA193984 standard; cDNA; 910 BP.
XX
AC AA193984;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 59.
XX
KM Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO20016719-A1.
XX
PD 13-SEP-2001.
XX
PE 02-MAR-2001; 2001WO-JP001629.
XX
PR 07-MAR-2000; 2000JP-00159195.
XX
PA (CHIB-) CHIBA PREFECTURE.

PA (HISM) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
PI
XX WPI; 2001-565584/63.
XX
XX Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human neuroblastoma,
PT malignancy and susceptibility indicator or tumor marker for anti-cancer
PT agents.
PS Claim 1; Page 79; 2979pp; Japanese.
XX
XX The invention relates to novel genes (AA193926-AA197963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumor markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes
XX
SQ Sequence 910 BP; 326 A; 176 C; 156 G; 215 T; 0 U; 37 Other;

Query Match 10.7%; Score 469; DB 4; Length 910;
Best Local Similarity 100.0%; Pred. No. 8.2e-158;
Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3867 TGGCCAGATTGAATGATGATTTCTGTTCTCCCTCCCTCCGAGGACCTTATTT 3926
DB 520 TGGCCAGATTGAATGATGATTTCTGTTCTCCCTCCCTCCGAGGACCTTATTT 461
QY 3927 TTAATGCCCCCTCTCTAGGTTAATTTCTCTTGAATTTGAATTTGAGAGAGGTTG 3986
DB 460 TTAATGCCCCCTCTCTAGGTTAATTTCTCTTGAATTTGAATTTGAGAGAGGTTG 401
QY 3987 GACAGTAGATTGCAAAAGTTCCAAAGTCAAAAGTCAAGTGTGAGTGTGGGGGAAA 4046
DB 400 GACAGTAGATTGCAAAAGTTCCAAAGTCAAAAGTCAAGTGTGAGTGTGGGGGAAA 341
QY 4047 ATTAGTCTTATTTTCCCTACATGGGATACACATGTTGAATTTCAATCTCACTGAAG 4106
DB 340 ATTAGTCTTATTTTCCCTACATGGGATACACATGTTGAATTTCAATCTCACTGAAG 281
QY 4107 CCGTCAGTTCTCTTAAGATAGTTGTTTCTTTTAAACAAAGTTAAGTATGT 4166
DB 280 CCGTCAGTTCTCTTAAGATAGTTGTTTCTTTTAAACAAAGTTAAGTATGT 221
QY 4167 TAATTAATTAATAAAATTCCTGTCGTCTACTTCAGCTTTGTTTATGCCATTTTCA 4226
DB 220 TAATTAATTAATAAAATTCCTGTCGTCTACTTCAGCTTTGTTTATGCCATTTTCA 161
QY 4227 ATTGTTGCTGTGTGTAATTTCAATTTTGAATTCATTTTGGATGATGATTTGGTT 4286
DB 160 ATTGTTGCTGTGTGTAATTTCAATTTTGAATTCATTTTGGATGATGATTTGGTT 101
QY 4287 GTCTTGAATATCTTATTAAGAGTTCAATTTGTAATTAATTAATTTGTTG 4335
DB 100 GTCTTGAATATCTTATTAAGAGTTCAATTTGTAATTAATTAATTTGTTG 52

RESULT 16
AA193983
ID AA193983 standard; cDNA; 904 BP.
XX
AC AA193983;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 58.
XX
KM Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX

PN WO200166719-A1.
XX 13-SEP-2001.
PD 02-MAR-2001; 2001WO-JP001629.
XX 07-MAR-2000; 2000JP-00159195.
XX (CHIB-) CHIBA PREFECTURE.
XX (HISM-) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
PI WPI; 2001-565584/63.
DR Nucleic acids originating in gene expressed in human neuroblastoma,
XX PT useful as probe or primer in diagnosing prognosis of human neuroblastoma,
XX PT malignancy and susceptibility indicator or tumor marker for anti-cancer
XX PT agents.
XX Claim 1; Page 78; 2979pp; Japanese.
XX The invention relates to novel genes (AA193926-AA197963) expressed in
XX human neuroblastoma. The nucleic acids are applicable as a probe or
XX primer in diagnosing the prognosis of human neuroblastoma, malignancy and
XX susceptibility indicators or tumour markers for anti-cancer agents. The
XX CC gene information for diagnosing prognosis is related to factors similar
XX to that for N-myc and TrkA genes
XX
SQ Sequence 904 BP; 265 A; 163 C; 198 G; 244 T; 0 U; 34 Other;

Query Match 10.3%; Score 449; DB 4; Length 904;
Best Local Similarity 100.0%; Pred. No. 1.2e-150;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1841 GCTACCTGGTGGATGAGGCGTACCGGATGAGATTCAGATCCAGAGAGCAAAACACC 1900
Db 52 GCTACCTGGTGGATGAGGCGTACCGGATGAGATTCAGATCCAGAGAGCAAAACACC 111

QY 1901 TCACCGGACGTGAGGAATAGCTGTGCTGGCAAGAACCTGTCTTCAGATAGTTGAGC 1960
Db 112 TCACCGGACGTGAGGAATAGCTGTGCTGGCAAGAACCTGTCTTCAGATAGTTGAGC 171

QY 1961 ATGCACATCCCGAGAGTGGCAAGACCTGTATATGTGACCTTTGTCTTCATATGTTAT 2020
Db 172 ATGCACATCCCGAGAGTGGCAAGACCTGTATATGTGACCTTTGTCTTCATATGTTAT 231

QY 2021 CACTGCTGATTAATACCTTCACTTCTTGACTTTGTTTTCATTACTGATTTTAC 2080
Db 232 CACTGCTGATTAATACCTTCACTTCTTGACTTTGTTTTCATTACTGATTTTAC 291

QY 2081 AAAAATCTTCACTTGGCTAATTTGTAGTTATGAGAGGATGATTTGATTTTCCCT 2140
Db 292 AAAAATCTTCACTTGGCTAATTTGTAGTTATGAGAGGATGATTTGATTTTCCCT 351

QY 2141 TTTTGGGAAATGAGGCTCTCAAGCTAAAGCTATAGATGAGCAATTCAGAGG 2200
Db 352 TTTTGGGAAATGAGGCTCTCAAGCTAAAGCTATAGATGAGCAATTCAGAGG 411

QY 2201 GTCTGTTTCATATGCTATGCTATGTTAAAGGGTAAAGGCTCTTTCATTAGATG 2260
Db 412 GTCTGTTTCATATGCTATGCTATGTTAAAGGGTAAAGGCTCTTTCATTAGATG 471

QY 2261 TGGAGATGAGGAGCCCTTCTTTAGA 2289
Db 472 TGGAGATGAGGAGCCCTTCTTTAGA 500

RESULT 17
AA194300/c
ID AA194300 standard; cDNA; 893 BP.
XX AC AA194300;

XX 13-NOV-2001 (first entry)
XX Human neuroblastoma expressed polynucleotide SEQ ID NO 375.
DE Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX Homo sapiens.
XX WO200166719-A1.
XX 13-SEP-2001.
XX 02-MAR-2001; 2001WO-JP001629.
XX 07-MAR-2000; 2000JP-00159195.
XX (CHIB-) CHIBA PREFECTURE.
XX (HISM-) HISAMITSU PHARM CO LTD.
XX Nakagawara A;
PI WPI; 2001-565584/63.
DR Nucleic acids originating in gene expressed in human neuroblastoma,
XX PT useful as probe or primer in diagnosing prognosis of human neuroblastoma,
XX PT malignancy and susceptibility indicator or tumor marker for anti-cancer
XX PT agents.
XX Claim 1; Page 318; 2979pp; Japanese.
XX The invention relates to novel genes (AA193926-AA197963) expressed in
XX human neuroblastoma. The nucleic acids are applicable as a probe or
XX primer in diagnosing the prognosis of human neuroblastoma, malignancy and
XX susceptibility indicators or tumour markers for anti-cancer agents. The
XX CC gene information for diagnosing prognosis is related to factors similar
XX to that for N-myc and TrkA genes
XX
SQ Sequence 893 BP; 323 A; 171 C; 162 G; 215 T; 0 U; 22 Other;

Query Match 9.8%; Score 427; DB 4; Length 893;
Best Local Similarity 100.0%; Pred. No. 8.5e-143;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3912 TGAGAACCTCTTAATTTATTTGTCCTCTTCAAGTTAAATCTCTTGATTTGACTTTG 3971
Db 468 TGAGAACCTCTTAATTTATTTGTCCTCTTCAAGTTAAATCTCTTGATTTGACTTTG 409

QY 3972 TTGAGAGAGAGGTTGACAGATTAAGCAAAAGTTCCAGTGCAGAAATTAACATGTTA 4031
Db 408 TTGAGAGAGAGGTTGACAGATTAAGCAAAAGTTCCAGTGCAGAAATTAACATGTTA 349

QY 4032 GAGTGTGGGGGAAATTAATGCTTATTTTCCCTACATGGGATACAACTGTGAATTC 4091
Db 348 GAGTGTGGGGGAAATTAATGCTTATTTTCCCTACATGGGATACAACTGTGAATTC 289

QY 4092 ATCTTCACTGAAGGCGCTGCACTTCTCTTAAACAATGTTGTTTCTTTAAACA 4151
Db 288 ATCTTCACTGAAGGCGCTGCACTTCTCTTAAACAATGTTGTTTCTTTAAACA 229

QY 4152 AGTTTAAGCTAGGTTTAATTAATTAATAAATAAATTTGCTGTCTCTTAATCACTTTGTT 4211
Db 228 AGTTTAAGCTAGGTTTAATTAATTAATAAATAAATTTGCTGTCTCTTAATCACTTTGTT 169

QY 4212 TTATGCCATTTGATTAATGTTGTCGTGTTGTAATTCATTAATTTGATTCGATTCGA 4271
Db 168 TTATGCCATTTGATTAATGTTGTCGTGTTGTAATTCATTAATTTGATTCGATTCGA 109

QY 4272 TGTGTAATAATGTTGCTCTGTAATAATCTTAAAGAGTTCAATTTGTAATAATTAATTAAT 4331
Db 108 TGTGTAATAATGTTGCTCTGTAATAATCTTAAAGAGTTCAATTTGTAATAATTAATTAAT 49

QY 4332 GTGGCTG 4338

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORF open reading frames 1 to 3161. The ORF
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiparasitic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiac; thrombolytic; coagulant; vasotonic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antihelmatic; antihypoid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORF-associated disorder. The nucleic acids can be used to express ORF
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers; proliferative disorders; neurodegenerative
CC disorders; osteoarthritis; graft vs host disease; cardiovascular disease,
CC diabetes mellitus; hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX

Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 0 U; 2 Other;

Query Match 8.7%; Score 379; DB 3; Length 2887;

Best Local Similarity 100.0%; Pred. No. 9.9e-126; Mismatches 0; Gaps 0;

Matches 379; Conservative 0; Indels 0; Gaps 0;

QY 1075 CCCCATTTGGAAATACAGAAAGAAAGACTCTGTTAAAGCTGCTCTGAATTTTCAA 1134
DB 2463 CCCCATTTGGAAATACAGAAAGAAAGACTCTGTTAAAGCTGCTCTGAATTTTCAA 2404
QY 1135 AAGTCAAAACAAAACTCTTAAGCACTCTCTATTAAAGAAACCTGTGTTCTTAT 1194
DB 2403 AAGTCAAAACAAAACTCTTAAGCACTCTCTATTAAAGAAACCTGTGTTCTTAT 2344
QY 1195 CTGAATCTGTTTAAACGTGAATTTGAGAGCCAAAGAAACCCCAAGAACCCCGGCTT 1254
DB 2343 CTGAATCTGTTTAAACGTGAATTTGAGAGCCAAAGAAACCCCAAGAACCCCGGCTT 2284
QY 1255 CAGTGAACACCCCAACCAAGACTCTCCACTCCCAAAAGAGACCCCAATCCCAAG 1314
DB 2283 CAGTGAACACCCCAACCAAGACTCTCCACTCCCAAAAGAGACCCCAATCCCAAG 2224
QY 1315 AGAAGCTTCTCAAGTAGAATGAAGATTGCGTACCTTTCCACCAAGAAATGTATT 1374
DB 2223 AGAAGCTTCTCAAGTAGAATGAAGATTGCGTACCTTTCCACCAAGAAATGTATT 2164
QY 1375 TGTGTGTTGGACACAGCTCTCCCATCACCCTTACCATTAAGGGAATCTCTCCAAAGA 1434
DB 2163 TGTGTGTTGGACACAGCTCTCCCATCACCCTTACCATTAAGGGAATCTCTCCAAAGA 2104
QY 1435 AGAGAGAGACTGTAGCAAG 1453
DB 2103 AGAGAGAGACTGTAGCAAG 2085

RESULT 20

ACH42852 standard; cDNA; 365 BP.

ACH42852;

13-OCT-2003 (first entry)

Human foetal liver/spleen cDNA #68.

Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;

genome mapping; biodiversity; genetic disorder.

Homo sapiens.

US2003073623-A1.

XX 17-APR-2003.

XX 30-JUL-2001; 2001US-00918995.

XX 30-JUL-2001; 2001US-00918995.

XX (DRMA/) DRMANAC R T.

XX (LABA/) LABAT I.

XX (STAC/) STACHE-CRAIN B.

XX (DICK/) DICKSON M C.

XX (JONE/) JONES L W.

XX Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

XX WPI; 2003-615964/58.

XX Claim 1; SEQ ID NO 30064; 44pp; English.

CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docID=20030073623
XX

Sequence 365 BP; 122 A; 86 C; 85 G; 72 T; 0 U; 0 Other;

Query Match 8.2%; Score 357; DB 9; Length 365;

Best Local Similarity 100.0%; Pred. No. 1e-117; Mismatches 0; Gaps 0;

Matches 357; Conservative 0; Indels 0; Gaps 0;

QY 919 AGCTGTTTCAGGGCTATGAAACTGAAGAGAGAGAAACAGAGCTATCTGAAGAAATTA 978
DB 9 AGCTGTTTCAGGGCTATGAAACTGAAGAGAGAGAAACAGAGCTATCTGAAGAAATTA 68
QY 979 AACTGAGTGCACGCGGAGCTTCCGAGACATCCCAAGACTGTGCTCCCAAGCCCTTCT 1038
DB 69 AACTGAGTGCACGCGGAGCTTCCGAGACATCCCAAGACTGTGCTCCCAAGCCCTTCT 128
QY 1039 CATGTGGCGGAGTGAAGAGGACATTAAGAAATCCCATTTGGAAGTACAGAAAGAA 1098
DB 129 CATGTGGCGGAGTGAAGAGGACATTAAGAAATCCCATTTGGAAGTACAGAAAGAA 188
QY 1099 AGACTCTGTTTAAAGAGCTGCTCTGAATTTTCAAAAGTCAAAACAAAACTCTTAAGC 1158
DB 189 AGACTCTGTTTAAAGAGCTGCTCTGAATTTTCAAAAGTCAAAACAAAACTCTTAAGC 248
QY 1159 ACTCTCTATTAAAGAGAAACCTGTGTTCTTATCTGAACCTGTTTGAAGCTGAAT 1218
DB 249 ACTCTCTATTAAAGAGAAACCTGTGTTCTTATCTGAACCTGTTTGAAGCTGAAT 308
QY 1219 TGAAGAGCCAAAGAAACCCCAAGAAAGCCCGGTCTTCAGTGACACCCCAAGAC 1275
DB 309 TGAAGAGCCAAAGAAACCCCAAGAAAGCCCGGTCTTCAGTGACACCCCAAGAC 365

Query Match	Best Local Similarity	Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1877 GGAGATCCAGAAAGCAACCACTCACCGACGCTGAGAAATAGCTGTGTGGCAGA	100.0%; Pred. No. 1.3e-99;	1936
117 GGAATCCAGAAAGCAACCACTCACCGACGCTGAGAAATAGCTGTGTGGCAGA		176
1937 ACCCTGTCTTCAGATAGTTGTAGCATGCCATTTCCGAGAGTGGCAGACCTGTATATGT		1996
177 ACCCTGTCTTCAGATAGTTGTAGCATGCCATTTCCGAGAGTGGCAGACCTGTATATGT		236
1997 GACCTTTGTCTTCACATATGTTATACCTCGCTGATATAACCTTTTCATACTCTTGTACT		2056
237 GACCTTTGTCTTCACATATGTTATACCTCGCTGATATAACCTTTTCATACTCTTGTACT		296
2057 TTGTTTTCATACCTGTGATTCACAAAAAAGCTTTTCATTCGGCTAAATGTGAGTTATGGA		2116
297 TTGTTTTCATACCTGTGATTCACAAAAAAGCTTTTCATTCGGCTAAATGTGAGTTATGGA		356
2117 GGGTGATTTGGGATTTCTTTTCCCTTTTGGGAAAATGGGCTCTCAAGCTAAAGCTATAGG		2176
357 GGGTGATTTGGGATTTCTTTTCCCTTTTGGGAAAATGGGCTCTCAAGCTAAAGCTATAGG		416
2177 ATGGCA 2182		
417 ATGGCA 422		

XX	AT22113	standard; cDNA to mRNA; 451 BP.
AC	AA22113;	
XX	02-AUG-1996	(first entry)
XX	Human gene signature	HUMGS03679.
XX	Gene signature; messenger RNA; mRNA; relative abundance; frequency;	
KW	human; cloning; mapping; non-biased library; diagnosis; detection;	
KW	cell typing; abnormal cell function; ss.	
XX	Homo sapiens.	
XX	W09514772-A1.	
XX	01-JUN-1995.	
XX	11-NOV-1994;	94WO-JP001916.
XX	12-NOV-1993;	93JP-00355504.
XX	(MATS/) MATSUBARA K.	
PA	(OKUB/) OKUBO K.	
XX	Matsubara K, Okubo K;	
XX	WPI; 1995-206931/27.	
XX	Single-stranded DNA for identifying gene signatures - isolated from 3' -	
PT	directed human cDNA library that reflects relative abundance of corresp.	
PT	mRNA in specific human tissues.	
XX	Claim 1; Page 1044; 2245bp; Japanese.	
XX	A single-stranded DNA (or its complementary strand or the corresp. double	
CC	-stranded DNA) which comprises one of the 7837 "GS" sequences given in	
CC	AA119001-126637 and which is able to hybridise to part of human genomic	
CC	DNA, cDNA or mRNA is claimed. The GS (Gene Signature) sequences were	
CC	obtained from 3'-directed cDNA libraries prepared from various human	
CC	tissues; synthesis of cDNA was initiated from the 3'-end of mRNA by using	
CC	poly(T) as the sole primer. Since the 3' - untranslated sequence is unique	
CC	to a particular mRNA species, almost all the 3'-oriented cDNAs hybridise	
CC	with specific mRNAs. Each library is constructed so as to reflect	
CC	accurately the relative abundance of different mRNAs in the particular	
CC	tissue from which it was derived. The appearance frequency of a given GS	
CC	in a cDNA library can be determined (esp. using primers and probes	
CC	derived from the GS sequences) as a means of diagnosing abnormal cell	
CC	function or for recognising different cell types	
XX	Sequence 451 BP; 105 A; 91 C; 93 G; 157 T; 0 U; 5 Other;	
XX	Query Match	6.5%; Score 284; DB 2; Length 451;
XX	Best Local Similarity	99.7%; Pred. No. 1.2e-91;
XX	Matches	334; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	3766	GGTATGAGTTAGCAAAATTTAACTATGTTGTTGTCCTTACCCAGGGAGCTCCCAAGTTT 38255
DB	51	GGATGTGAGTTAGCAAAATTTAACTATGTTGTTGTCCTTACCCAGGGAGCTCCCAAGTTT 110
QY	3826	CTGACTTGAAGTACAGTGAAGAAATCAAGAGTGTCATCTGGCCAGATTAGTAGAT 38855
DB	111	CTGACTTGAAGTACAGTGAAGAAATCAAGAGTGTCATCTGGCCAGATTAGTAGAT 170
QY	3886	TCTATTTCTTGTTGTTCCCTCTCCCTGAGGACCTTATTTTATTTATGTCCTCTTCTAG 39445
DB	171	TCTATTTCTTGTTGTTCCCTCTCCCTGAGGACCTTATTTTATTTATGTCCTCTTCTAG 230
QY	3946	GTTAATTTCTCTTGATTGATTGACTTTGTTGAGAGAGAGTGGACAGTAGATTACCAAGT 40055

Db 231 GTTAATTCCTTGAATTGACTTTGTTGGAAGAGAGTTGACAGTAAATTAGCAAGT 290
QY 4006 TCCTAAGTCGCAAAATTAACAGTGTGTAGAGTGTGGGGGAAAAATTAAGTCTTATTTTCCCT 4065
Db 291 TCCAAATGTGCAAAATTAACAGTGTGTAGAGTGTGGGGGAAAAATTAAGTCTTATTTTCCCT 350
QY 4066 ACATGGATACACACTGTGAATTCATCTTCAAC 4100
Db 351 ACATGGATACACACTGTGAATTCATCTTCAAC 385

RESULT 23
ID ADF59738 standard; cDNA; 2106 BP.
XX ADF59738;
AC ADF59738;
XX
DT 12-FEB-2004 (first entry)
XX
DE Human contig polynucleotide sequence SEQ ID NO:2105.
XX
KM biological activity; genetic engineering; hybridisation probe; oligomer;
KM primer; chromosome mapping; gene mapping; recombinant protein production;
XX human; gene; ss.
XX Homo sapiens.
XX
XX MO2003080795-A2.
XX
XX 02-OCT-2003.
XX
PF 09-AUG-2002; 2002MO-US025485.
XX
PR 09-AUG-2001; 2001US-0311261P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Yang Y, Wang Z, Weng G, Ma Y;
XX
DR WPI; 2003-876918/81.
XX
P-P-SDB; ADF60190.
XX
PT New polynucleotides, useful as hybridization probes, oligomers or
PT primers, for chromosome or gene mapping, for the recombinant production
PT of proteins, and for generating antisense DNA or RNA.
XX
XX
PS Example 2; SEQ ID NO 2105; 571bp; English.
XX
CC The present invention describes isolated polynucleotide sequences (I),
CC which encode polypeptides (II) with biological activity. Also described:
CC (1) a vector comprising (I); (2) an expression vector comprising (I); (3)
CC a host cell genetically engineered to comprise (I) which is operatively
CC associated with a regulatory sequence that modulates expression of (I) in
CC the host cell; (4) a polypeptide (II) encoded by (I); (5) a composition
CC comprising the polypeptide of (4) and a carrier; (6) an antibody directed
CC against the polypeptide of (4); (7) detecting (I) or the polypeptide of
CC (4) in a sample; (8) identifying a compound that binds to the polypeptide
CC of (4); (9) producing the polypeptide of (4); and (10) a collection of
CC polynucleotides comprising at least one of the polynucleotide sequences
CC (I). The polynucleotides (I) can be used as hybridisation probes,
CC oligomers or primers, for chromosome or gene mapping, for the recombinant
CC production of proteins, and for generating antisense DNA or RNA. The
CC present sequence represents a human contig polynucleotide sequence, which
CC is used in an example from the present invention.
XX
XX
SQ Sequence 2106 BP; 526 A; 569 C; 588 G; 422 T; 0 U; 1 Other;

Query Match 6.2%; Score 270; DB 10; Length 2106;
Best Local Similarity 99.3%; Pred. No. 9.5e-87;
Matches 540; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 313 GGGGCTTGTACTCCCGCGGGCGGGCCCCCGGCGAGAGAGAGAGCTGGGGCGGTT 372
|||||

Db 217 GGGGCTTGTACTCCCGCGGGCGGGCCCCCGGCGAGAGAGAGAGCTGGGGCGGTT 276
QY 373 CGGTGCTCTGCTCGGTCTCGGCTCCGCGCCCGGCGCAACAGAGAGCGGCTATTTGGGGGAGGCTTG 432
Db 277 CGGTGCTCTGCTCGGTCTCGGCTCCGCGCCCGGCGCAACAGAGAGCGGCTATTTGGGGGAGGCTTG 336
QY 433 CGGACGCGGAGCGGCTCGAGCGGCTCGGCGCCCGGCGCAAGATACAGGCGGTGCGCATTCAGA 492
Db 337 CGGACGCGGAGCGGCTCGAGCGGCTCGGCGCCCGGCGCAAGATACAGGCGGTGCGCATTCAGA 396
QY 493 CGGCTCTCTCGGTGCGGCGGCGGCGCAAGAGCTTACGCTGAGCTGAGGCGCAAGAGGTTGGG 552
Db 397 CGGCTCTCTCGGTGCGGCGGCGGCGGCGCAAGAGCTTACGCTGAGGCGCAAGAGGTTGGG 456
QY 553 CGGCTCTCTCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 612
Db 457 CGGACTCCCGCGCTGCGCAACGCTCGGAGCCGCGGCGGAGCCCGCACCACTTCTGCGCG 516
QY 613 GGGCGGCAACCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 672
Db 517 GGGCGGCAACCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 576
QY 673 GGAAGATAGAGAGAGAGCGCTCTCGGAGGT-GGTGCGGCTCGGAGGCTCGCACTAG 721
Db 577 GGAAGATAGAGAGAGAGCGCTCTCGGAGGTGCGGCTCGGAGGCTCGCACTAG 636
QY 732 GCGGCTGCTCTCAACAGATCTTCTGCTGCAATTGAACTCATTCAGACGACGACG 791
Db 637 GCGGCTGCTCTCAACAGATCTTCTGCTGCAATTGAACTCATTCAGACGACGACG 696
QY 792 CAGCTCAGGCGCAAGAGAGAGATCGAGAGCTGAAGTCAAGAGAGACGCTCCCT 851
Db 697 CAGCTCAGGCGCAAGAGAGAGATCGAGAGCTGAAGTCAAGAGAGACGCTCCCT 756
QY 852 GCTC 855
Db 757 GCTC 760

RESULT 24
ABV88816/c
ID ABV88816 standard; cDNA; 230 BP.
XX
XX ABV88816;
AC
XX
DT 13-DEC-2002 (first entry)
XX
DE Human colon cancer related cDNA SEQ ID NO 2127.
XX
XX
KM Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
KM ss.
XX
OS Homo sapiens.
XX
PN MO200258534-A2.
XX
PD 01-AUG-2002.
XX
PF 16-NOV-2001; 2001MO-US043704.
XX
PR 20-NOV-2000; 2000US-0252222P.
PR 06-FEB-2001; 2001US-0267011P.
PR 28-MAR-2001; 2001US-0279670P.
PR 10-JUL-2001; 2001US-0304037P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;
XX
DR WPI; 2002-608400/65.
XX
PT New isolated tumor colon polynucleotide and polypeptide, useful for the
PT diagnosis, prevention and/or treatment of cancer, in particular colon

PT cancer.
PS Claim 1; SEQ ID NO 2127; 266pp + Sequence Listing; English.
XX
CC The invention relates to a human colon tumour expressed polynucleotide
CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (I) any of
CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (II)
CC complements of (I); (III) at least 20 contiguous residues of (I); (IV)
CC sequences that hybridize to (I), under moderately stringent conditions;
CC (V) sequences having at least 75% or 90% identity to (I); or (VI)
CC degenerate variants of (I). The compositions and methods of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC cancer, particularly colon cancer. (I) can be used in gene therapy and
CC (II) and (III) are useful in pharmaceutical compositions such as vaccines.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://www.wipo.int/pub/published_pct_sequences)
XX
SQ Sequence 230 BP; 84 A; 46 C; 43 G; 57 T; 0 U; 0 Other;
XX
Query Match 5.3%; Score 230; DB 6; Length 230;
Best Local Similarity 100.0%; Pred. No. 2.8e-72;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 3870 CCAGATTTAAGTATGATTTCTTCTGTTCTCCCTCCCTGAGGACCTTATTTTA 3929
Db 230 CCAGATTTAAGTATGATTTCTTCTGTTCTCCCTCCCTGAGGACCTTATTTTA 171
XX
QY 3930 TTGTCCCTCTTCTAGTATTTCTCTTTGATTTGACTTTGTTGAGAGAGGTTGAC 3989
Db 170 TTGTCCCTCTTCTAGTATTTCTCTTTGATTTGACTTTGTTGAGAGAGGTTGAC 111
XX
QY 3990 AGRAGATTAACAAGTCCCAAGTCAAAATTAAGCTGTGTAGAGTGGGGGAAATTT 4049
Db 110 AGRAGATTAACAAGTCCCAAGTCAAAATTAAGCTGTGTAGAGTGGGGGAAATTT 51
XX
QY 4050 AGCTTATTTTCTCTCATGAGGATACACACTGTGATTCATCTTCAA 4099
Db 50 AGCTTATTTTCTCTCATGAGGATACACACTGTGATTCATCTTCAA 1
XX
RESULT 25
ACH46937 standard; cDNA; 411 BP.
XX
AC ACH46937;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human infant brain cDNA #1000.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Dmanac RT, Labat-I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.

XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 34149; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12769-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20030073623
XX
SQ Sequence 411 BP; 125 A; 74 C; 94 G; 118 T; 0 U; 0 Other;
XX
Query Match 4.9%; Score 213; DB 9; Length 411;
Best Local Similarity 99.5%; Pred. No. 3e-66;
Matches 383; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
XX
QY 2179 GGCAAGATTCAGAACTTTCAGGGGCTCTTCTATACATTTGCCATGTTAAAGGGGTAA 2238
Db 1 GGCAAGATTCAGAACTTTCAGGGGCTCTTCTATACATTTGCCATGTTAAAGGGGTAA 60
XX
QY 2239 AGGGCTCTCTTCAATTAACATGTGAAGATGAAGCAGCCCTTCTTAGAGCTGTGCT 2298
Db 61 AGGGCTCTCTTCAATTAACATGTGAAGATGAAGCAGCCCTTCTTAGAGCTGTGCT 120
XX
QY 2299 GCATGGCACTCTTCTCACCCCTGTGACACCCCTCTTATAGTGGTATAGATTTTAAAC 2358
Db 121 GCATGGCACTCTTCTCACCCCTGTGACACCCCTCTTATAGTGGTATAGATTTTAAAC 180
XX
QY 2359 CTAATAATTAACAACAACCTCAACCATG-AGCTTAGACACCAAGAGGAATGCAAGTG 2417
Db 181 CTAATAATTAACAACAACCTCAACCATGAGCTTTAGACCAAGAAAGAAATGCAAGTG 240
XX
QY 2418 AAGGATGAAGCAAGCCATCTTCAAGAGTAGAAGAAACATCGAGAGTGTGATTAAC 2477
Db 241 AAGGATGAAGCAAGCCATCTTCAAGAGTAGAAGAAACATCGAGAGTGTGATTAAC 300
XX
QY 2478 TGTCTGAAAAGATAGTGTTCATTGAACTATTTCTGTGATACAGTCATGTGGAGGGA 2537
Db 301 TGTCTGAAAAGATAGTGTTCATTGAACTATTTCTGTGATACAGTCATGTGGAGGGA 360
XX
QY 2538 TGTTCGCTGTGATTAATTTTTCAG 2562
Db 361 TGTTCGCTGTGATTAATTTTTCAG 385
XX
RESULT 26
ABK39516 standard; cDNA; 542 BP.
XX
ID ABK39516
XX
AC ABK39516;
XX
DT 21-MAY-2002 (first entry)
XX
DE DNA encoding lung tumour protein associated contig 228.
XX
KW Lung tumour; cancer; T cell; immune response stimulator; cyostatic;
KW contig; ss.

XX	Homo sapiens.
XS	
XN	WO200204514-A2.
XX	
PX	17-JAN-2002.
PD	
PF	10-JUL-2001; 2001WO-US022058.
XX	
PR	11-JUL-2000; 2000US-00614124.
PR	29-AUG-2000; 2000US-00651563.
PR	08-SEP-2000; 2000US-00658824.
PR	26-SEP-2000; 2000US-00671135.
PR	06-OCT-2000; 2000US-00677419.
PR	30-OCT-2000; 2000US-00702705.
PR	13-DEC-2000; 2000US-00736457.
PR	03-MAY-2001; 2001US-00849626.
PA	(CORI-) CORIXA CORP.
XX	
PI	Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW,
PI	Mammetakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS, Mcnabb A;
PI	Wang A, Fanger N, Switzer A, Monell PD, Clapper JD;
XX	
PS	WPI; 2002-164634/21.
DR	
PT	Novel polynucleotide encoding a lung tumor polypeptide useful for stimulating and/or expanding T cells specific for a tumor protein.
CC	
CS	Claim 1; SEQ ID NO 1554; 223pp; English.
XX	
CC	The invention describes an isolated polynucleotide and polypeptide useful for stimulating and/or expanding T cells specific for a tumour protein for determining the presence of a cancer in a patient. A composition containing the polynucleotide and/or polypeptide is useful for treating a lung cancer in a patient. The polypeptide is useful for removing tumour cells from a biological sample. The polynucleotide is also useful as probe or primer to detect the level of mRNA encoding a tumour protein. This sequence represents a contig of a DNA sequence encoding a lung tumour associated protein, described in the method of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
CC	
CX	
SQ	Sequence 542 BP; 125 A; 104 C; 119 G; 188 T; 0 U; 6 Other;
	Query Match 3.5%; Score 152; DB 6; Length 542; Best Local Similarity 98.9%; Pred. No. 1.8e+44;
	Matches 352; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
DY	3395 TGATGATGTGTGTGTGTGTGTGGCATAGCGTTTCATTGTGAACCTCATTGCTTAGAG 3454 Db TGTGTGTGTGTGTGTGTGTGTGTGTGGCATAGCGTTTCATTGTGAACCTCATTGCTTAGARAG 125
DY	3455 AGTGCGCTCTCTATAAGGAACTGCCTGTAAAATTCAATGCCAGAAGATGTAGAGAA 3514 Db 126 AGTGCGCTCTCTATAAGGAACTGCCTGTAAAACTTATGGACAAGATGTAGAGAA 185
DY	3515 ATAGACTTAATTCACATGAGGGGCTCMTATCACCACCTTAAGAGAGATTTCTAGAA 3574 Db ATAGACTTAATTCACATGAGGGGCTCMTATCACCACCTTAAGAGAGATTTCTAGAAA 245
DY	3575 AACGGGCGCAATTTTTCTTGTTCTCCATCAATTTAAATGTGGCAGGCTGTCAAGTTTCT 3633 Db AACGGGCGCAATTTTTCTTGTTCTCCATCAATTTAAATGTGGCAGGCTGTCAAGTTTCT 305
DY	3635 TAATCTTAACCTATGTGATATTTTCTTCGTACGCTGCCAAAAAGAAAAAGCCCCAATCAG 3694 Db TAATCTTAACCTATGTGATATTTTCTTCGTACGCTGCCAAAAAGAAAAAGCCCCAATCAG 365
DY	3695 TGTCTCTGCATTTTGTCTTTGATCCCTCAGTTTCTTCTTGATTTCAAGATGTGTC 3750 Db TGTCTCTGCATTTTGTCTTTGATCCCTCAGTTTCTTCTTGATTTCAAGATGTGTC 421

RESULT 27
 ACAL1845
 ID ACAL1845 standard; cDNA, 542 BP.
 XX
 AC ACAL1845;
 XX
 DT 05-JUN-2003 (first entry)
 DE Human lung cancer library SCL3/4, cDNA SEQ ID 1554.
 XX
 XX Human; lung cancer; ss; lung tumour; cytostatic; vaccine;
 KW T cell expansion; CD4; CD8.
 XX
 OS Homo sapiens.
 XX
 PN US2002197669-A1.
 XX
 PD 26-DEC-2002.
 XX
 PF 03-MAY-2001; 2001US-00849626.
 XX
 PR 13-DEC-2000; 2000US-00736457.
 PA (BANG/) BANGUR C S.
 PA (FANG/) FANGER G R.
 PA (WANG/) WANG A.
 PA (MANG/) MANG T.
 PA (SWIT/) SWITZER A P.
 PA (MCNE/) MCNEILL P D.
 PA (CLAP/) CLAPPER J D.
 PI Bangur CS, Fanger GR, Wang A, Wang T, Switzer AP, Mcneill PD,
 PI Clapper JD;
 XX
 DR WPI; 2003-352750/33.
 PT Novel lung cancer polynucleotide encoding lung cancer protein, useful for
 detecting the presence of lung cancer in a patient, and in pharmaceutical
 compositions, e.g. vaccines, for treating lung cancer.
 XX
 PS Example 1; Page; 72pp; English.
 XX
 CC The invention relates to a polynucleotide encoding a lung tumour protein,
 CC comprising a sequence selected from any of the 14 sequences mentioned in
 CC the specification, or a sequence (S2) mentioned in specification,
 CC complement of S1, sequences consisting of at least 20 contiguous residues
 CC of S1, sequences that hybridise to S1, sequences having 75%, preferably
 CC 90%, identity to S1, or degenerate variants of S1. Also included are an
 CC isolated polypeptide (comprising a sequence (S3) selected from any one of
 CC the 4 amino acid sequences mentioned in the specification, a sequence
 CC encoded by the polynucleotide, or sequences having at least 70%,
 CC preferably 90%, identity to a sequence encoded by the polynucleotide), an
 CC expression vector comprising the polynucleotide operably linked to an
 CC expression control sequence, a host cell transformed or transfected with
 CC the vector, an isolated antibody (or its antigen-binding fragment) that
 CC specifically binds to the polypeptide, detecting the presence of a cancer
 CC in a patient, a fusion protein comprising the polypeptide, an
 CC oligonucleotide that hybridises to S1 under moderately stringent
 CC conditions, stimulating and/or expanding T cells specific for a tumour
 CC protein (comprising contacting T cells with the polynucleotide, protein
 CC or antigen-presenting cells, under conditions and for a time sufficient
 CC to permit the stimulation and/or expansion of T cells) and inhibiting the
 CC development of a cancer in a patient (by incubating CD4⁺ and/or CD8⁺ T
 CC cells isolated from a patient with the polynucleotide, protein or antigen
 CC presenting cells that express the polynucleotide, such that T cells
 CC proliferate, administering to the patient an effective amount of the
 CC proliferated T cells, and thus inhibiting the development of a cancer in
 CC the patient. The polynucleotide, protein and cells are useful in a
 CC composition for stimulating an immune response in a patient, and for
 CC treating a cancer in a patient (particularly lung cancer). The
 CC oligonucleotide is useful for determining the presence of a cancer in a

DE Human prostate expression marker cDNA 56437.
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX Homo sapiens.
OS
XX
PN W0200160860-A2.
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001MO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
PS Claim 1; Page 10890; 11750BP; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 211 BP; 89 A; 37 C; 29 G; 55 T; 0 U; 1 Other;
Query Match 3.1%; Score 134; DB 5; Length 211;
Best Local Similarity 100.0%; Pred. No. 5.8e-38;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4239 GTTGTAATTCATACCTTTGATACCATTTCTGATGCTGTAATAATGCTGCTTGTAAATA 4238
DB 52 GTTGTAATTCATACCTTTGATACCATTTCTGATGCTGTAATAATGCTGCTTGTAAATA 111
QY 4299 TCTATTAAGAGTCAATGTAATAAATTAATGCTGCTGTTAAAAAATGCTGCTTGTAAATA 4358
DB 112 TCTTATTAAGAGTCAATGTAATAAATTAATGCTGCTGTTAAAAAATGCTGCTTGTAAATA 171
QY 4359 AAAAAAAAAAAAAA 4372
DB 172 AAAAAAAAAAAAAA 185
RESULT 32
ABK62818
XX ABK62818 standard; cDNA; 615 BP.
XX
AC ABK62818;
XX
DT 18-JUN-2002 (first entry)
XX
DE Rat sequence differentially expressed in response to a hepatotoxin #725.

XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
KM differential expression; centrillobular necrosis; steatosis.
XX
XX Rattus norvegicus.
OS
XX
PN W0200210453-A2.
PD 07-FEB-2002.
XX
PF 30-JUL-2001; 2001MO-US023872.
XX
PR 31-JUL-2000; 2000US-0222040P.
PR 02-NOV-2000; 2000US-0244880P.
PR 11-MAY-2001; 2001US-0290029P.
PR 15-MAY-2001; 2001US-0290645P.
PR 22-MAY-2001; 2001US-0292336P.
PR 06-JUN-2001; 2001US-0295788P.
PR 13-JUN-2001; 2001US-0297457P.
PR 19-JUN-2001; 2001US-0298884P.
PR 09-JUL-2001; 2001US-0303459P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
DR WPI; 2002-241625/29.
XX
PT Predicting toxic effects of compounds or the progression of these toxic
PT effects by determining the changes in gene expression in tissues or cells
PT exposed to the toxin and comparing these to gene expression in unexposed
PT tissues or cells.
PS Claim 1; SEQ ID NO 725; 239BP; English.
XX
XX The invention relates to methods for predicting toxic effects of
CC compounds or the progression of these toxic effects by determining the
CC global changes in gene expression in tissues or cells exposed to the
CC toxin and comparing these to gene expression in unexposed tissues or
CC cells. Also included are methods of predicting at least one toxic effect
CC of a compound or progression of a toxic effect, preferably the
CC hepatotoxicity of a compound, comprising detecting the level of
CC expression in a tissue or cell sample exposed to the compound of two or
CC more genes listed in the specification, where differential expression of
CC the genes is indicative of at least one toxic effect or progression. The
CC method can also be used to identify an agent which modulates the toxic
CC response and predict cellular pathways that a compound modulates in a
CC cell. The methods utilize a set of at least two probes (on a solid
CC support in kit form), where each of the probes comprises a sequence that
CC specifically hybridizes to a gene listed in the specification, a computer
CC system comprising a database containing information identifying the
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC set of genes comprising at least two genes listed in the specification,
CC and a user interface to view the information used to present information
CC identifying the expression level in a tissue or cell of at least one gene
CC listed in the specification. The method is useful for elucidating global
CC changes in gene expression and for identifying toxicity markers in
CC tissues or cell exposed to a known toxin. The genes may be used as
CC toxicity markers in drug screening and toxicity assays. The genes and
CC gene expression information may be used as diagnostic markers for the
CC prediction or identification of the physiological state of tissue or cell
CC sample that has been exposed to a compound or agent. Hepatotoxicity is
CC characterized by centrillobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent
XX
SQ Sequence 615 BP; 84 A; 235 C; 224 G; 71 T; 0 U; 1 Other;
Query Match 2.5%; Score 110; DB 6; Length 615;
Best Local Similarity 100.0%; Pred. No. 1.8e-29;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 447 GCGTCAGCCCGCCGCAAGTATCAGCGCGTGTGCTCCATTGACAGCGGCTCTTCGTG 506

DB 339 GGCTGAGCCCCCGGCCCAAGTATCAAGCGGCTCTGCCATTAGAGAGGGCTCTTCGTG 398
QY 507 GCGGCGGCCCAAGAGCCTTACGCCCCCTGCGCTGGGACACAGGCTGGGCGGC 556
DB 399 GCGGCGGCCCAAGAGCCTTACGCCCCCTGCGCTGGGACACAGGCTGGGCGGC 448

RESULT 33

AAC25273

ID AAC25273 standard; cDNA; 317 BP.

AC AAC25273;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 29348.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-00200610.

PR 26-FEB-1999; 99US-0122487P.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS Claim 1; SEQ ID NO 29348; 71bp + Sequence listing; English.

CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
CC
XX

SQ Sequence 317 BP; 58 A; 58 C; 67 G; 115 T; 0 U; 9 Other;

Query Match 2.4%; Score 106; DB 3; Length 317;
Best Local Similarity 100.0%; Pred. No. 5.5e-28;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3569 TAGAAAAAAGCGGCGAGATTTCTTGTCTCCATCATTTTAATGTGGAGGCTGTTCAG 3628

DB 212 TAGAAAAAAGCGGCGAGATTTCTTGTCTCCATCATTTTAATGTGGAGGCTGTTCAG 271

QY 3629 TTTTCTTACTCTTACCTTANGATATTTCTTCTGTAAGTGTCCAA 3674

DB 272 TTTTCTTACTCTTACCTTANGATATTTCTTCTGTAAGTGTCCAA 317

RESULT 34

ID ACA03031 standard; cDNA; 482 BP.

AC ACA03031;

DT 22-MAY-2003 (first entry)

DE Lung cancer therapy diagnosis associated cDNA #1520.

XX Lung cancer; cytostatic; vaccine; gene therapy; cancer; gene; ss.

OS Homo sapiens.

PN US2002172952-A1.

PD 21-NOV-2002.

PF 10-JUL-2001; 2001US-00902941.

PR 30-JUN-1999; 99US-00346492.

PR 15-OCT-1999; 99US-00419356.

PR 17-DEC-1999; 99US-00466867.

PR 30-DEC-1999; 99US-00476300.

PR 06-MAR-2000; 2000US-00519642.

PR 22-MAR-2000; 2000US-00533077.

PR 10-APR-2000; 2000US-00546259.

PR 27-APR-2000; 2000US-00560406.

PR 05-JUN-2000; 2000US-00589184.

PR 11-JUL-2000; 2000US-00614124.

PR 29-AUG-2000; 2000US-00651563.

PR 08-SEP-2000; 2000US-00658824.

PR 26-SEP-2000; 2000US-00671325.

PR 06-OCT-2000; 2000US-00677419.

PR 30-OCT-2000; 2000US-00702705.

PR 13-DEC-2000; 2000US-00736457.

PR 03-MAY-2001; 2001US-00849626.

PA (CORI-) CORIXA CORP.

PI Henderson RA, Wang T, Watanabe Y, Johnson JC, Retter MW;

PI Durham M, Carter D, Fanger GR, Vedvick TS, Bangur CS, McNabb A;

DR WPI; 2003-328427/31.

XX New polynucleotide, useful for preparing a composition for treating or
PT inhibiting development of cancer, e.g. lung cancer.
XX

PS Example 1; SEQ ID NO 1554; 82bp; English.

XX The invention describes an isolated polynucleotide comprising one of 32
CC sequences, complement or degenerate variants of them. The polynucleotide
CC is useful for preparing a composition e.g. a vaccine or for gene therapy,
CC for treating or inhibiting development of cancer, e.g. lung cancer. This
CC sequence represents a polynucleotide associated with the compositions and
CC methods for the therapy and diagnosis of lung cancer
CC
XX

SQ Sequence 482 BP; 107 A; 94 C; 102 G; 174 T; 0 U; 5 Other;

Query Match 2.3%; Score 101; DB 8; Length 482;
Best Local Similarity 100.0%; Pred. No. 3.1e-26;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3650 GATATTTCTTCGTACGTGTCCAAAAAGAAAGAACCCCAATCAGTGTCTTGACTTTG 3709

DB 261 GATATTTCTTCGTACGTGTCCAAAAAGAAAGAACCCCAATCAGTGTCTTGACTTTG 320

QY 3710 TTCTTGATCCCTCAGTGTCTTTCTTGAATTCACAGTGTGC 3750

DB 321 TTCTTGATCCCTCAGTGTCTTTCTTGAATTCACAGTGTGC 361

RESULT 35

ABX52399
ID ABX52399 standard; cDNA; 346 BP.
XX
XX AC ABX52399;
XX
XX 25-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #2328.
XX
XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
KM muscle deposition; fat deposition; genome mapping; gene identification;
KM gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX
XX US2002137160-A1.
XX
XX 26-SEP-2002.
XX
XX 26-OCT-2001; 2001US-00983965.
PF 26-DEC-1998; 98US-0113678P.
PR 15-DEC-1999; 99US-00465231.
XX
XX (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
XX WPI; 2003-102386/09.
DR
XX
XX
XX Purified nucleic acid molecules, useful for genome mapping, gene
PT identification and analysis, cattle breeding or preparation of constructs
PT for cattle gene expression and genetically improved cattle.
XX
XX
XX Claim 2; SEQ ID NO 2328; 38pp; English.
XX
XX The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMPD), derived from
CC cattle, and the LMPD nucleic acid can specifically hybridise to a second
CC nucleic acid molecule comprising any of 5912 nucleotide sequences,
CC appearing as ABX50072-ABX55983, or complements of them. Also included are
CC ; (1) a transformed cell having a nucleic acid comprising an LMPD nucleic
CC acid linked to a promoter and a 3' non-translated sequence that
CC functions in the cell to cause termination of transcription and addition
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC (2) determining a level or pattern of a molecule in a bovine cell or
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC of the 5912 nucleic acid sequences or its complement or fragment) with a
CC complementary nucleic acid molecule obtained from the bovine cell or
CC tissue, where hybridisation between the marker nucleic acid and the
CC complementary nucleic acid permits the detection of the molecule; and (b)
CC detecting the level or pattern of the complementary nucleic acid, where
CC the detection of the complementary nucleic acid is predictive of the
CC level or pattern of the molecule. The LMPD nucleic acid is used for
CC determining a level or pattern of a molecule in a bovine cell or tissue.
CC It is useful for genome mapping, gene identification and analysis, cattle
CC breeding, preparation of constructs for use in cattle gene expression, or
CC for genetically improving cattle. The present sequence is one of the 5912
CC bovine LMPD EST (expressed sequence tag) nucleic acids. Note: The present
CC sequence was not shown in the specification but was obtained in
CC electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?docID=20020137160
XX
XX Sequence 346 BP; 84 A; 69 C; 85 G; 108 T; 0 U; 0 Other;
XX
XX Query Match 1.4%; Score 63; DB 8; Length 346;
XX Best Local Similarity 100.0%; Pred. No. 1,3e-12;
XX Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX

Db	62	TGCTGTGTGTGTGGCTATGGGTTTTCATTGTGTACTGCATCTTCTTAGAGAGTGGGCT	121
Qy	3463	CTC	3465
Db	122	CTC	124
RESULT 36			
ABN49572			
ID	ABN49572	standard; DNA; 60 BP.	
XX	ABN49572;		
AC	15-JUL-2002	(first entry)	
XX			
XX	Human spliced transcript detection oligonucleotide SEQ ID NO:22320.		
DE			
XX	Human; mouse; rat; splice transcript; detection; RNA transcript;		
KW	splice variant; transcriptome; oligonucleotide library; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	NO200210449-A2.		
XX			
PD	07-FEB-2002.		
XX			
PF	20-JUL-2001; 2001WO-1B001903.		
XX			
PR	28-JUL-2000; 2000US-0221607P.		
FR	02-MAY-2001; 2001US-0287724P.		
XX			
XX	(COMP-) COMPUGEN INC.		
PA			
XX			
PI	Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;		
XX	WPI; 2002-257393/30.		
XX			
PT	New oligonucleotide libraries comprising oligonucleotides which		
PT	selectively hybridize to mRNAs transcribed from a transcription unit of a		
PT	genome, useful for detecting tissue-, pathology-, and developmental-		
PT	specific genes.		
XX			
XX			
PS	Example 1; SEQ ID NO 22320; 47bp; English.		
XX			
CC	The present invention describes oligonucleotide libraries for detecting		
CC	messenger RNAs that populate a (sub-)transcriptome, where the (sub-		
CC)transcriptome comprises messenger RNAs transcribed from multiple		
CC	transcription units that populate a genome. The library comprises several		
CC	oligonucleotides, each capable of hybridizing selectively to a set of		
CC	messenger RNAs transcribed from a given transcription unit of the genome,		
CC	which encodes one or more messenger RNA splice variants. The		
CC	oligonucleotide libraries are useful for detecting mRNAs from a		
CC	biological sample, in expression profiling studies, in qualitatively or		
CC	quantitatively characterizing the corresponding transcriptome, and in		
CC	detecting RNA transcripts and splice variants of human or animal		
CC	transcriptomes. The libraries may also be used as specialized mini		
CC	libraries to detect transcripts of a sub-transcriptome under a particular		
CC	biological or pathological state, and so allowing the detection of tissue		
CC	- and pathology-specific genes such as those genes only expressed in		
CC	specific tissue under a specific pathological condition; to detect		
CC	developmental specific genes; and to detect RNA transcripts and splice		
CC	variants of a transcriptome of a patient suffering from a particular		
CC	disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from		
CC	rats, humans and mice, which are used in the exemplification of the		
CC	present invention. N.B. The sequence data for this patent did not form		
CC	part of the printed specification, but was obtained in electronic format		
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences		
XX			
SO	Sequence 60 BP; 18 A; 14 C; 16 G; 12 T; 0 U; 0 Other;		
Query Match	1.4%;	Score 60;	DB 6; Length 60;
Best Local Similarity	100.0%;	Pred. No. 2.1e-11;	

	Matches	60;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1201	CTGTTTGTAAACGTGAATTAGAGAACCCAGAAGAAACCCCGTGCTTCATGTGG	1260							
Dd	1	CTGTTTGTAAACGTGAATTAGAGAACCAAGAAACCCAGAAAAGCCCCGGTCTTCATGTGG	60							
	RESULT 37									
ID	ABK62109/c									
XX	ABK62109 standard; cDNA; 590 BP.									
AC	ABK62109;									
XX	18-JUN-2002 (first entry)									
DE	Rat sequence differentially expressed in response to a hepatotoxin #16.									
XX										
KW	Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;									
XX	differential expression; centrilobular necrosis; steatosis.									
OS	Rattus norvegicus.									
PN	MO200210453-A2.									
PD	07-FEB-2002.									
PF	30-JUL-2001; 2001WO-US023872.									
PR	31-JUL-2000; 2000US-0222040P.									
PR	02-NOV-2000; 2000US-0244880P.									
PR	11-MAY-2001; 2001US-0290029P.									
PR	15-MAY-2001; 2001US-0290645P.									
PR	22-MAY-2001; 2001US-0292336P.									
PR	06-JUN-2001; 2001US-0295798P.									
PR	13-JUN-2001; 2001US-0297457P.									
PR	19-JUN-2001; 2001US-0298884P.									
PA	09-JUL-2001; 2001US-0303459P.									
(GENE-)	GENE LOGIC INC.									
PI	Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;									
DR	WPI; 2002-241625/29.									
PT	Predicting toxic effects of compounds or the progression of these toxic									
PT	effects by determining the changes in gene expression in tissues or cells									
PT	exposed to the toxin and comparing these to gene expression in unexposed									
PT	tissues or cells.									
PS	Claim 1; SEQ ID NO 16; 239pp; English.									
XX	The invention relates to methods for predicting toxic effects of									
CC	compounds or the progression of these toxic effects by determining the									
CC	global changes in gene expression in tissues or cells exposed to the									
CC	toxin and comparing these to gene expression in unexposed tissues or									
CC	cells. Also included are methods of predicting at least one toxic effect									
CC	of a compound or progression of a toxic effect, preferably the									
CC	hepatotoxicity of a compound, comprising detecting the level of									
CC	expression in a tissue or cell sample exposed to the compound of two or									
CC	more genes listed in the specification, where differential expression of									
CC	the genes is indicative of at least one toxic effect or progression. The									
CC	method can also be used to identify an agent which modulates the toxic									
CC	response and predict cellular pathways that a compound modulates in a									
CC	cell. The method utilise a set of at least two probes (on a solid									
CC	support in kit form), where each of the probes comprises a sequence that									
CC	specifically hybridises to a gene listed in the specification, a computer									
CC	system comprising a database containing information identifying the									
CC	expression level in a tissue or cell sample exposed to a hepatotoxin of a									
CC	set of genes comprising at least two genes listed in the specification,									
CC	and a user interface to view the information used to present information									
CC	identifying the expression level in a tissue or cell of at least one gene									
CC	listed in the specification. The method is useful for elucidating global									
CC	changes in gene expression and for identifying toxicity markers in									

CC	tissues or cell exposed to a known toxin. The genes may be used as
CC	toxicity markers in drug screening and toxicity assays. The genes and
CC	gene expression information may be used as diagnostic markers for the
CC	prediction or identification of the physiological state of tissue or cell
CC	sample that has been exposed to a compound or agent. Hepatotoxicity is
CC	characterised by centrilobular necrosis and steatosis. The present
CC	sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC	which is differentially expressed in response to a hepatotoxic agent
XX	
SQ	Sequence 590 BP; 205 A; 134 C; 106 G; 143 T; 0 U; 2 Other;
	Query Match 1.2%; Score 51; DB 6; Length 599;
	Best Local Similarity 100.0%; Pred. No. 2.3e-08;
	Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
Oy	4286 TGCTTGTGAATATCTTTATAAGAGTCCAATTGTAATAACTATTGTGGC 4336
Dd	55 TGCTTGTAATATCTTTATAAGAGTCCAATTGTAATAACTATTGTGGC 5
RESULT 38	
ID	ADB55014/c
XX	ADB55014 standard; DNA; 590 BP.
AC	ADB55014;
DT	04-DEC-2003 (first entry)
XX	
DE	Toxicity-related gene, SEQ ID 40.
KM	Toxic; toxin; gene expression profile; hepatotoxicity; liver;
OS	drug screening; toxicity assay; ds.
XX	Unidentified.
PN	WO2003064624-A2.
PD	07-AUG-2003.
PF	31-JAN-2003; 2003WO-US003194.
PR	31-JAN-2002; 2002US-00060087.
PR	15-MAR-2002; 2002US-0364045P.
PR	15-MAR-2002; 2002US-0364055P.
PR	30-DEC-2002; 2002US-0436644P.
PA	(GENE-) GENE LOGIC INC.
P1	Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M,
DR	WPI, 2003-689530/65.
XX	
PT	Predicting a toxic effect of a compound, useful in identifying toxicity
PT	markers in liver tissues or cells for drug screening and toxicity assays,
PT	comprises preparing gene expression profile of tissue or cells exposed to
PT	the compound.
PS	Claim 1; SEQ ID NO 40; 1156bp; English.
XX	
CC	The present invention relates to a method for predicting a toxic effect
CC	of a compound. The method comprises preparing a gene expression profile
CC	of a tissue or cell sample exposed to the compound, and comparing the
CC	gene expression profile to a database comprising SEQ ID 1-14925, where
CC	differential expression of the gene indicates at least one toxic effect.
CC	The method is useful for predicting at least one toxic effect of a
CC	compound, predicting hepatotoxicity or the progression of a toxic effect
CC	of a compound, identifying an agent that modulates the onset or
CC	progression of a toxic response, predicting the cellular pathways that a
CC	compound modulates in a cell, and identifying an agent that modulates at
CC	least one activity of a protein. The method and compositions of the
CC	present invention using a database of genes having liver toxin-induced
CC	differential expression, are useful in identifying toxicity markers in
CC	liver tissues or cells for drug screening and toxicity assays. Note: The

CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.

SO Sequence 590 BP, 205 A, 134 C, 106 G, 143 T, 0 U, 2 Other;

Query Match 1.2%; Score 51; DB 10; Length 590;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4286 TGCCTGTAATATCTTATTAAGGTCATGTAATGAATGCTGAC 4336

DB 55 TGCTTGTAAATATCTTATTAAGGTCATGTAATGAATGCTGAC 5

RESULT 39

AAH36707 standard; cDNA; 196 BP.

AAH36707;

03-SEP-2001 (first entry)

Human colon cancer antigen encoding cDNA SEQ ID NO:3789.

Human; colon cancer; colon cancer antigen; diagnosis; detection;
colorectal carcinoma; ss.

Homo sapiens.

WO200122920-A2.

05-APR-2001.

28-SEP-2000; 2000MO-US026524.

29-SEP-1999; 99US-0157137P.

03-NOV-1999; 99US-0163280P.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Barash SC, Birse CE, Rosen CA;

WPI; 2001-235357/24.

P-PDSB; AAG77300.

Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
useful for preventing, diagnosing and/or treating colorectal cancers.

Claim 1; Page 5645-5647; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
cancer-associated nucleic acid molecules (N) and proteins (P), where the
proteins are collectively known as colon cancer antigens. The colon
cancer antigens have cytostatic activity and can be used in gene therapy
and vaccine production. N and P may be used in the prevention, diagnosis
and treatment of diseases associated with inappropriate P expression. For
example, N and P may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of P by expressing inactive proteins or to
supplement the patient's own production of P. Additionally, N may be used
to produce the colon cancer-associated Ps, by inserting the nucleic acids
into a host cell and culturing the cell to express the proteins. N and P
can be used in the prevention, diagnosis and treatment of colorectal
carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
sequences used in the exemplification of the present invention. N.B.
Pages 666 to 682 and page 7053 of the sequence listing were missing at
time of publication, meaning no sequences are present for SEQ ID NO:1027
to 1052, 7921 and 7922

Sequence 196 BP, 39 A, 42 C, 42 G, 62 T, 0 U, 11 Other;

Query Match 1.1%; Score 49; DB 4; Length 196;

Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3884 ATTCATTTCTGTTCTTCCCTCCCTGAGGACCTTATTTATTTG 3932

DB 79 ATTCATTTCTGTTCTTCCCTCCCTGAGGACCTTATTTATTTG 127

RESULT 40

AAH1178 standard; cDNA; 50 BP.

AAH1178;

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 15253.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
gene therapy; chromosome mapping; ss.

Homo sapiens.

EPI033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-00200610.

26-FEB-1999; 99US-0122487P.

(GEST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
diagnostic, forensic, gene therapy and chromosome mapping procedures.

Claim 1; SEQ ID NO 15253; 71pp + Sequence Listing; English.

The present sequence is one of a large number of 5' ESTs derived from
cDNAs encoding secreted proteins. No ORF has yet been conclusively
identified within the present sequence. The 5' ESTs were prepared from
total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
sequences usually correspond mainly to the 3' untranslated region (UTR)
of the mRNA because they are often obtained from oligo-dT primed cDNA
libraries. Such ESTs are not well suited for isolating cDNA sequences
derived from the 5' ends of mRNAs and even in those cases where longer
cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
ESTs are derived from mRNAs with intact 5' ends and can therefore be used
to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
diagnostic, forensic, gene therapy and chromosome mapping procedures.
They are used to obtain upstream regulatory sequences and to design
expression and secretion vectors

Sequence 50 BP, 17 A, 13 C, 12 G, 8 T, 0 U, 0 Other;

Query Match 1.1%; Score 47; DB 3; Length 50;

Best Local Similarity 100.0%; Pred. No. 9.4e-07;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 AAACACCTACCGACGCTGAGAAATAGCTGTGCTGCAAGACCC 1940

DB 1 AAACACCTACCGACGCTGAGAAATAGCTGTGCTGCAAGACCC 47

RESULT 41

AAAT47669 standard; cDNA; 827 BP.

AAAT47669

AAAT47669

AAAT47669

AAAT47669

AAAT47669

AAAT47669

AAAT47669

AAAT47669

```
AC AAT47669;
XX
XX 19-MAY-1997 (first entry)
XX
XX Human G protein gamma-2 subunit cDNA.
DE
XX G protein; signal transduction; agonist; antagonist; diagnosis; therapy;
KM ss.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH CDS 232..448
FT /*tag= a
FT
XX
XX WO9637513-A1.
XX
XX 28-NOV-1996.
XX
XX 22-MAY-1995; 95WO-US006406.
XX
XX 22-MAY-1995; 95WO-US006406.
XX
XX 22-MAY-1995; 95WO-US006406.
XX
XX (SMK ) SMITHKLINE BEECHAM CORP.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (WEIS-) WEIS CENT RES.
XX
XX Robishaw JD, Kunsch CA;
XX
XX MPI: 1997-021140/02.
XX P-PSDB; AAM09412.
XX
XX New nucleic acid encoding gamma sub:units of human G protein - used to
XX detect mutation(s) in, or altered levels of, the sub:units and to screen
XX for agonists and antagonists of G protein interaction.
XX
XX Claim 1; Page 32-33; 40pp; English.
XX
XX cDNA clones (AAT47669-75) respectively code for human G protein gamma-2,
XX gamma-3, gamma-4, gamma-5, gamma-7, gamma-10 and gamma-11 subunits
XX (AAM09412-18). Several human cDNA libraries were partially sequenced to
XX identify expressed sequence tags (EST). By matching sequences of EST to
XX genes of known structure, the 7 new subunit sequences were isolated. A
XX mixture of cDNA clones for gamma-2, -3, -5 and -7 has been deposited as
XX ATCC 97137. The isolated cDNA clones can be used to produce recombinant
XX gamma subunits in prokaryotic or eukaryotic host cells for use in the
XX development of therapeutic and diagnostic agents. The nucleic acids can
XX also be used to produce probes to detect mutations in human G protein
XX subunit sequences and for chromosome identification
XX
XX Sequence 827 BP; 239 A; 209 C; 202 G; 177 T; 0 U; 0 Other;
SQ
Query Match 0.9%; Score 38; DB 2; Length 827;
Best Local Similarity 100.0%; Pred. No. 0.00094;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4335 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 789 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 826
RESULT 42
ABX14564
ID ABX14564 standard; cDNA; 827 BP.
XX
XX ABX14564;
AC
XX
XX 13-MAR-2003 (first entry)
XX
XX Human gamma2 subunit cDNA.
DE
XX
XX Human; gamma2 subunit; radiolabelling; gene; ss.
XX
XX
```

```
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 233..448
FT /*tag= a
FT /product= "Human gamma2 subunit"
FT
XX
XX US2002106678-A1.
XX
XX 08-AUG-2002.
XX
XX 22-OCT-2001; 2001US-00982809.
XX
XX 30-JUN-1998; 98US-00952772.
XX
XX (ROBI/) ROBISHAW J D.
XX (KUNS/) KUNSCH C A.
XX
XX Robishaw JD, Kunsch CA;
XX
XX MPI: 2002-706221/76.
XX P-PSDB; ABG73049.
XX
XX New nucleic acids encoding the G protein gamma subunit family are useful
XX in research for the discovery of treatments and diagnostics for human
XX disease.
XX
XX Disclosure; Page 14; 17pp; English.
XX
XX The invention relates to the human gamma2, gamma3, gamma4, gamma5,
XX gamma7, gamma10 and gamma11 subunit polypeptides and the polynucleotides
XX encoding them. Mutated forms of human gamma subunits can be detected in a
XX patient by obtaining genomic DNA from the patient, hybridizing it to
XX radiolabelled gamma2, gamma3, gamma4, gamma5, gamma7, gamma10 or gamma11
XX cRNA or antisense DNA sequences, and distinguishing matched from
XX mismatched sequences to detect mutated forms having point mutations.
XX Altered levels of gamma2, gamma3, gamma4, gamma5, gamma7, gamma10 or
XX gamma11 subunits in various tissues of a host can be detected by
XX incubating a sample tissue on a solid support so that proteins in the
XX sample bind to the support, incubating an antibody specific to the gamma
XX subunit with the support and detecting the bound antibody. The gamma
XX subunits are useful in research for the discovery of treatments and
XX diagnostics for human disease. This sequence represents cDNA encoding the
XX human gamma2 subunit
XX
XX Sequence 827 BP; 239 A; 209 C; 202 G; 177 T; 0 U; 0 Other;
SQ
Query Match 0.9%; Score 38; DB 6; Length 827;
Best Local Similarity 100.0%; Pred. No. 0.00094;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4335 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 789 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 826
RESULT 43
AAF18137
ID AAF18137 standard; DNA; 1654 BP.
XX
XX AAF18137;
AC
XX
XX 14-MAR-2001 (first entry)
XX
XX Lung cancer associated polynucleotide sequence SEQ ID 156.
XX
XX Human; lung cancer associated protein; neuroprotective; cytoprotective;
XX cardioactive; immunomodulatory; muscular active; vulnerary;
XX gastrointestinal; nephroprotective; antiinfective; gynecological;
XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX proliferative disorder; wound healing; infectious disease; ds.
XX
XX Homo sapiens.
OS
```



```

XX PN WO200055180-A2.
XX XX 21-SEP-2000.
XX XX
XX PF 08-MAR-2000; 2000MO-US005918.
XX PR 12-MAR-1999; 99US-0124270P.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX PI Ruben SM;
XX DR MPI; 2000-587514/55.
XX DR P-PSDB; AAB58261.
XX XX
XX PT Lung cancer associated gene sequences, referred to as lung cancer
XX PT anti-gens, useful for treatment, prevention, and diagnosis of disorders
XX PT such as lung cancer.
XX PS Claim 1; Page 620-621; 1425pp; English.
XX XX
XX CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
XX CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX CC associated proteins and polynucleotide sequences, their agonists, and
XX CC antagonists may have neuroprotective; cytostatic; cardiostatic;
XX CC immunomodulatory; muscular active general; vulnary; gastrointestinal
XX CC general; nephrotoxic; antineoplastic; gynecological; or antibacterial
XX CC activity. The invention also includes antibodies specific for the protein
XX CC or polynucleotide sequences. The lung cancer associated polynucleotide
XX CC sequences may be used for detection of lung cancer, chromosome
XX CC identification, as chromosome markers, and for numerous other diagnostic
XX CC or research purposes. The proteins may be used to treat disorders such as
XX CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
XX CC cardiovascular, renal, and proliferative disorders. The proteins may also
XX CC be used in the treatment of wounds and infectious diseases.
XX CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
XX CC used in the course of the invention for the identification and
XX CC characterization of the polynucleotide and protein sequences
XX XX
XX SQ Sequence 1654 BP; 354 A; 505 C; 501 G; 290 T; 0 U; 4 Other;
XX XX
XX Query Match 0.9%; Score 38; DB 3; Length 1654;
XX Best Local Similarity 100.0%; Pred. No. 0.00083;
XX Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX
XX QY 4335 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
XX Db 1605 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1642
XX XX
XX RESULT 44
XX ABX47809/c
XX ID ABX47809 standard; cDNA; 241 BP.
XX XX
XX AC ABX47809;
XX XX
XX DT 21-FEB-2003 (first entry)
XX XX
XX DE Bovine EST associated with lactation/muscle/fat deposition #12974.
XX XX
XX XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
XX KM muscle deposition; fat deposition; genome mapping; gene identification;
XX KM gene analysis; cattle breeding.
XX XX
XX OS Bos Taurus.
XX XX
XX PN US2002137139-A1.
XX PD 26-SEP-2002.
XX XX
XX PF 24-SEP-2001; 2001US-00960352.

```

```

XX XX
XX PR 12-JAN-1999; 99US-0115707P.
XX PR 11-JAN-2000; 2000US-00480902.
XX XX
XX PA (BYAT/) BYATT J C.
XX PA (MATH/) MATHIALAGAN N.
XX PA (TAON/) TAO N.
XX PA (WARR/) WARREN W C.
XX XX
XX PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX XX
XX DR MPI; 2003-110599/10.
XX XX
XX PT New nucleic acid associated with lactation, and muscle and fat
XX PT deposition, useful for genome mapping, gene identification and analysis,
XX PT cattle breeding, or for genetically improving cattle.
XX PS Claim 2; SEQ ID NO 12974; 245pp; English.
XX XX
XX CC The invention relates to a purified nucleic acid molecule associated with
XX CC lactation or muscle and fat deposition (designated LMPD), derived from
XX CC cattle, and the LMPD nucleic acid can specifically hybridize to a second
XX CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
XX CC appearing as ABX34836-ABX49947, or complements of them. Also included are
XX CC ; (1) a transformed cell having a nucleic acid comprising an LMPD nucleic
XX CC acid linked to a promoter and a 3' non-translated sequence that
XX CC functions in the cell to cause termination of transcription and addition
XX CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
XX CC (2) determining a level or pattern of a molecule in a bovine cell or
XX CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
XX CC of the 15112 nucleic acid sequences or its complement or fragment) with a
XX CC complementary nucleic acid molecule obtained from the bovine cell or
XX CC tissue, where hybridisation between the marker nucleic acid and the
XX CC complementary nucleic acid permits the detection of the molecule; and (b)
XX CC detecting the level or pattern of the complementary nucleic acid, where
XX CC the detection of the complementary nucleic acid is predictive of the
XX CC level or pattern of the molecule. The LMPD nucleic acid is used for
XX CC determining a level or pattern of a molecule in a bovine cell or tissue.
XX CC It is useful for genome mapping, gene identification and analysis, cattle
XX CC breeding, preparation of constructs for use in cattle gene expression, or
XX CC for genetically improving cattle. The present sequence is one of the
XX CC 15112 bovine LMPD EST (expressed sequence tag) nucleic acids. Note: The
XX CC present sequence was not shown in the specification but was obtained in
XX CC electronic format from the USPTO web site:
XX CC seqdata.uspto.gov/sequence.html?docid=20020137139
XX XX
XX SQ Sequence 241 BP; 55 A; 55 C; 35 G; 96 T; 0 U; 0 Other;
XX XX
XX Query Match 0.8%; Score 37; DB 8; Length 241;
XX Best Local Similarity 100.0%; Pred. No. 0.0027;
XX Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX
XX QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
XX Db 107 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 71
XX XX
XX RESULT 45
XX ABV56639
XX ID ABV56639 standard; cDNA; 243 BP.
XX XX
XX AC ABV56639;
XX XX
XX DT 17-SEP-2002 (first entry)
XX XX
XX DE Human prostate expression marker cDNA 56630.
XX XX
XX KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KM pharmacogenomic marker; gene; ss.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200160860-A2.

```

```
XX
PD 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US005171.
XX
XX 17-FEB-2000; 2000US-0183119P.
XX 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JR,
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 10921; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABY00010-ABY62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 243 BP; 132 A; 28 C; 35 G; 48 T; 0 U; 0 Other;
SQ
Query Match 0.8%; Score 37; DB 5; Length 243;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 155 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 191
RESULT 46
AA183077
ID AA183077 standard; cDNA; 403 BP.
XX
XX AA183077;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human polynucleotide SEQ ID NO 3137.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
OS
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US004927.
XX
XX 28-FEB-2000; 2000US-00515126.
XX 18-MAY-2000; 2000US-00577409.
PR
```

```
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX P-PsDB; AA003146.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 1; SEQ ID NO 3137; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at http://wipo.int/pub/published\_pct\_sequences
XX
XX Sequence 403 BP; 192 A; 46 C; 82 G; 83 T; 0 U; 0 Other;
SQ
Query Match 0.8%; Score 37; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 92 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 128
RESULT 47
AA183503
ID AA183503 standard; cDNA; 410 BP.
XX
XX AA183503;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human polynucleotide SEQ ID NO 3563.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
OS
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US004927.
XX
XX 28-FEB-2000; 2000US-00515126.
XX 18-MAY-2000; 2000US-00577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX P-PsDB; AA003572.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
```

XX Claim 1; SEQ ID NO 3563; 1399bp + Sequence Listing; English.
PS
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 410 BP; 206 A; 64 C; 62 G; 77 T; 0 U; 1 Other;
XX
Query Match 0.8%; Score 37; DB 4; Length 410;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 91 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 127
RESULT 48
ACH21614
ID ACH21614 standard; cDNA; 410 BP.
XX
AC ACH21614;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult liver cDNA #1226.
XX
KW Human, ss; sequencing by hybridisation; SHH; expressed sequence tag; EST;
KM genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 8826; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH2789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences

CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20030073623
XX
SQ Sequence 410 BP; 206 A; 64 C; 62 G; 77 T; 0 U; 1 Other;
XX
Query Match 0.8%; Score 37; DB 9; Length 410;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 91 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 127
RESULT 49
ACH19590
ID ACH19590 standard; cDNA; 418 BP.
XX
AC ACH19590;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult lung cDNA #593.
XX
KW Human, ss; sequencing by hybridisation; SHH; expressed sequence tag; EST;
KM genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 6802; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH2789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide

CC sequences are also useful as hybridisation probes, as oligomers for PCR,
 CC for chromosome and gene mapping, in the recombinant production of
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide
 CC is useful for generating antibodies specific for it. The present sequence
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docId=20030073623
 CC
 SQ Sequence 418 BP, 141 A, 81 C, 86 G, 99 T, 0 U, 11 Other;
 Query Match 0.8%; Score 37; DB 9; Length 418;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4336 CTGTTAAAAA
 DB 57 CTGTTAAAAA
 RESULT 50
 AAI87359
 ID AAI87359 standard; cDNA; 432 BP.
 XX
 AC AAI87359;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 7419.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US004927.
 XX
 PR 28-FEB-2000; 2000US-00515126.
 XX
 PR 18-MAY-2000; 2000US-00577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI, 2001-514838/56.
 XX
 DR P-PDB; AA007428.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukaemia, inflammation and immune disorders.
 XX
 PS Claim 1; SEQ ID NO 7419; 1399bp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 432 BP, 212 A, 46 C, 90 G, 83 T, 0 U, 1 Other;

Query Match 0.8%; Score 37; DB 4; Length 432;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4336 CTGTTAAAAA
 DB 92 CTGTTAAAAA
 RESULT 51
 AAL08202/c
 ID AAL08202 standard; cDNA; 446 BP.
 XX
 AC AAL08202;
 XX
 DT 07-DEC-2001 (first entry)
 XX
 DE Human breast cancer expressed polynucleotide 659.
 XX
 KW Human; breast cancer; cell marker; cytostatic; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200151628-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 10-JAN-2001; 2001WO-US000798.
 XX
 PR 14-JAN-2000; 2000US-0176077P.
 PR 14-MAR-2000; 2000US-0189167P.
 PR 24-MAR-2000; 2000US-0192099P.
 PR 29-MAR-2000; 2000US-0193480P.
 PR 15-MAY-2000; 2000US-0205230P.
 PR 09-JUN-2000; 2000US-0211315P.
 PR 25-JUL-2000; 2000US-0220534P.
 XX
 PA (MIL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Lillie J, Xu Y, Wang Y, Steinmann K;
 XX
 DR WPI, 2001-451856/48.
 XX
 PT New peptide useful as a marker for the diagnosis of breast cancer.
 XX
 PS Claim 1; Page 197-198; 3695pp; English.
 XX
 CC The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity
 XX
 SQ Sequence 446 BP, 117 A, 85 C, 121 G, 118 T, 0 U, 5 Other;
 Query Match 0.8%; Score 37; DB 4; Length 446;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4336 CTGTTAAAAA
 DB 87 CTGTTAAAAA
 RESULT 52
 ACH23288
 ID ACH23288 standard; cDNA; 450 BP.
 XX

AC ACH23288;
XX
XX 13-OCT-2003 (first entry)
XX
XX Human adult ovary cDNA #1668.
DE
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KM genome mapping; biodiversity; genetic disorder.
XX
XX Homo sapiens.
XX US2003073623-A1.
XX
XX 17-APR-2003.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX (STAC/) STACHE-CRAIN B.
XX (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX
XX Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
PI WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 10500; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
XX 38043 cDNA sequences, appearing as ACH2789-ACH50831, whose sequence was
XX determined by the technique of SBH (sequencing by hybridisation). Also
XX included is a purified polypeptide comprising a sequence corresponding to
XX a reading frame of the novel polynucleotide. The nucleic acid sequences
XX are useful in diagnostics as expressed sequence tags (EST) for
XX identifying expressed genes or for physical mapping of the human genome,
XX in forensic, in assessing biodiversity, or in identifying mutations
XX responsible for genetic disorders and other traits. The nucleotide
XX sequences are also useful as hybridisation probes, as oligomers for PCR,
XX for chromosome and gene mapping, in the recombinant production of
XX protein, or in generating antisense DNA or RNA. The purified polypeptide
XX is useful for generating antibodies specific for it. The present sequence
XX is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?docID=20030073623
XX
XX Sequence 450 BP; 150 A; 105 C; 80 G; 91 T; 0 U; 24 Other;
SQ
XX
XX Query Match 0.8%; Score 37; DB 9; Length 450;
XX Best Local Similarity 100.0%; Pred. No. 0.0024;
XX Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 339 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 375

RESULT 53
ID AAQ90525 standard; cDNA; 458 BP.
XX
XX AAQ90525;
XX
XX 25-MAR-2003 (revised)
DT 02-NOV-1995 (first entry)

XX
XX Rat SIII 15 kDa subunit cDNA.
DE
XX RNA-polymerase transcription factor SIII; ss.
KM
XX
XX Rattus sp.
OS
XX
XX Key Location/Qualifiers
FT CDS 46..384
FT /*tag= a
XX
XX EP655498-A1.
XX
XX 31-MAY-1995.
XX
XX 29-NOV-1994; 94EP-00250286.
XX
XX 30-NOV-1993; 93US-00160087.
XX
XX (OKLA-) OKLAHOMA MED RES FOUND.
XX
XX Conaway RC, Conaway JM, Bradsher JN;
PI WPI; 1995-195587/26.
XX
XX P-PSDB; AAR75086.
XX
XX Novel RNA polymerase transcription factor SIII - useful for modulating
PT the rate of transcription by RNA polymerase II.
XX
XX Claim 38; Page 18; 43pp; English.
XX
XX Tryptic peptides (given in AAR75090-93) of purified rat SIII p15 subunit
XX were used to design primers (AAQ90531-32) for the PCR amplification of
XX cDNA from a rat lambda-gt11 library. cDNAs encoding the complete p15
XX protein (AAR75086) were obtained by screening rat liver and brain lambda
XX ZAP II cDNA libraries. 2 overlapping clones were combined to generate the
XX sequence given in AAQ90525. (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 458 BP; 186 A; 73 C; 85 G; 114 T; 0 U; 0 Other;
SQ
XX
XX Query Match 0.8%; Score 37; DB 2; Length 458;
XX Best Local Similarity 100.0%; Pred. No. 0.0024;
XX Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 405 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 441

RESULT 54
ID AAT59993 standard; cDNA; 458 BP.
XX
XX AAT59993;
XX
XX 08-JUN-1997 (first entry)
DT
XX Rat RNA polymerase transcription factor elongin C subunit cDNA.
DE
XX RNA polymerase transcription factor; elongation factor; elongin C; SIII;
KM ss; de.
XX
XX Rattus sp.
OS
XX
XX Key Location/Qualifiers
FT CDS 46..384
FT /*tag= a
XX
XX MO9709426-A1.
XX
XX 13-MAR-1997.
PD
XX 09-SEP-1996; 96MO-US014522.
PF

CC cell growth, neurological conditions, and diseases related to the
 CC pituitary gland, colon, breast, lungs and prostate. These disorders
 CC include Alzheimer's, Parkinson, diabetes, dwarfism, colour blindness,
 CC retinal pigmentosa, asthma, depression, schizophrenia, sleeplessness,
 CC hypertension, anxiety, stress and renal failure. This sequence represents
 CC novel human G protein coupled receptor HGPBMY18 3' untranslated region
 CC (UTR)
 CC
 SO Sequence 632 BP; 219 A; 113 C; 134 G; 166 T; 0 U; 0 Other;
 Query Match 0.8%; Score 37; DB 8; Length 632;
 Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 4336 CTGTTAAAAA
 Db 582 CTGTTAAAAA
 RESULT 57
 ADB47644
 ID ADB47644 standard; cDNA; 632 BP.
 AC ADB47644;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Human cDNA encoding GPCR, HGPBMY18, 3' UTR.
 XX
 KW Human; ss; gene; G protein-coupled receptor; GPCR; endocrine disorder;
 KW pituitary disorder; growth hormone; prolactin; luteinising hormone;
 KW follicle-stimulating hormone; thyroid-stimulating hormone;
 KW adrenocorticotrophin; vasopressin; oxytocin; aberrant growth;
 KW aberrant lactation; aberrant sexual characteristic development;
 KW testosterone; oestrogen; aberrant water homeostasis; hypogonadism;
 KW Addison's disease; hypothyroidism; Cushing's disease; agromegaly;
 KW gigantism; lethargy; osteoporosis; aberrant calcium homeostasis;
 KW aberrant potassium homeostasis; reproductive disorder;
 KW developmental disorder; colon cancer; breast cancer; prostate cancer;
 KW lung cancer; untranslated region.
 KW
 XX Homo sapiens.
 OS
 XX
 PN US2003129653-A1.
 XX
 PD 10-JUL-2003.
 XX
 PF 30-SEP-2002; 2002US-00262313.
 XX
 PR 16-JAN-2001; 2001US-0261782P.
 PR 27-JUL-2001; 2001US-0308540P.
 PR 14-NOV-2001; 2001US-00992331.
 XX
 PA (FEDE/) FEDER J N.
 PA (MINT/) MINTIER G.
 PA (RAMA/) RAMANATHAN C S.
 PA (HAWK/) HAWKEN D R.
 XX
 PI Feder JN, Mintier G, Ramanathan CS, Hawken DR;
 XX
 DR WPI; 2003-678603/64.
 XX
 PT New human G-protein coupled receptor, HGPBMY18, useful for treating and
 PT diagnosing disease such as cancer, hypothyroidism, Cushing's disease and
 PT osteoporosis.
 XX
 PS Disclosure; Fig 4; 71pp; English.
 XX
 CC The invention relate to an isolated human G protein-coupled receptor
 CC (GPCR), HGPBMY18, polynucleotide encoding a novel human GPCR HGPBMY18.
 CC Also included are expression vectors, host cells, a fusion protein
 CC comprising HGPBMY18 and an Fc portion of a human immunoglobulin protein,
 CC an anti-HGPBMY18 antibody its antigenic epitope, screening a library of

CC molecules or compounds with a polynucleotide to identify at least one
 CC molecule or compound which specifically binds to the polynucleotide
 CC sequence and screening for candidate compounds capable of modulating
 CC activity of a G-protein coupled receptor-encoding polypeptide. HGPBMY18
 CC or its antagonist (small molecule, peptide, and antisense molecule) is
 CC useful for treating a disease, disorder, or condition related to the
 CC endocrine, gastrointestinal, reproductive, pulmonary, or neural system.
 CC Diseases include endocrine disorders, disorders of the pituitary,
 CC aberrant growth hormone synthesis and/or secretion, aberrant prolactin
 CC synthesis and/or secretion, aberrant luteinising hormone synthesis and/or
 CC secretion, aberrant follicle-stimulating hormone synthesis and/or
 CC secretion, aberrant thyroid-stimulating hormone synthesis and/or
 CC secretion, aberrant adrenocorticotrophin synthesis and/or secretion,
 CC aberrant vasopressin secretion, aberrant oxytocin secretion, aberrant
 CC growth, aberrant lactation, aberrant sexual characteristic development,
 CC aberrant testosterone synthesis and/or secretion, aberrant oestrogen
 CC synthesis and/or secretion, aberrant water homeostasis, hypogonadism,
 CC Addison's disease, hypothyroidism, Cushing's disease, agromegaly,
 CC gigantism, lethargy, osteoporosis, aberrant calcium homeostasis, aberrant
 CC potassium homeostasis, reproductive disorders, developmental disorders,
 CC colon cancer, related proliferative condition of the colon, breast
 CC cancer, related proliferative condition of the breast, prostate cancer,
 CC related proliferative condition of the prostate, lung cancer, and related
 CC proliferative condition of the lung. In addition determining the presence
 CC or amount of expression of HGPBMY18 is useful for diagnosing a
 CC (susceptibility to a) pathological condition such as colon cancer,
 CC related proliferative condition of the colon, breast cancer, related
 CC proliferative condition of the breast, prostate cancer, related
 CC proliferative condition of the prostate, lung cancer, and related
 CC proliferative condition of the lung. The present sequence represents the
 CC 5' or 3' UTR of the cDNA encoding HGPBMY18.
 CC
 SO Sequence 632 BP; 219 A; 113 C; 134 G; 166 T; 0 U; 0 Other;
 Query Match 0.8%; Score 37; DB 9; Length 632;
 Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 4336 CTGTTAAAAA
 Db 582 CTGTTAAAAA
 RESULT 58
 ADR40538
 ID ADR40538 standard; DNA; 632 BP.
 AC ADR40538;
 XX
 DT 04-NOV-2004 (first entry)
 XX
 DE Human orphan G-protein coupled receptor HGPBMY18 3'-UTR.
 XX
 KW human; G-coupled receptor; HGPBMY18; endocrine disorder;
 KW gastrointestinal disorder; reproductive disorder; pulmonary disorder;
 KW neural-related disorder; pituitary; growth hormone; prolactin;
 KW luteinising hormone; follicle-stimulating hormone;
 KW thyroid-stimulating hormone; adrenocorticotrophin; vasopressin; oxytocin;
 KW lactation; sexual characteristic; testosterone; oestrogen;
 KW water homeostasis; hypogonadism; Addison's disease; hypothyroidism;
 KW Cushing's disease; agromegaly; gigantism; lethargy; osteoporosis;
 KW calcium homeostasis; potassium homeostasis; colon cancer; breast cancer;
 KW prostate cancer; lung cancer; renal tumour; kidney tumour; ds.
 KW
 XX Homo sapiens.
 OS
 XX
 PN US2004161823-A1.
 XX
 PD 19-AUG-2004.
 XX
 PF 30-JAN-2004; 2004US-00768878.
 XX
 PR 14-NOV-2000; 2000US-0248483P.

```
PR 16-JAN-2001; 2001US-0261782P.
PR 27-JUL-2001; 2001US-0308540P.
PR 14-NOV-2001; 2001US-0092331.
PR 30-SEP-2002; 2002US-00262313.
XX
XX (FEDER/) FEDER J N.
PA (MINT/) MINTIER G.
PA (RAMA/) RAMANATHAN C S.
PA (HAWK/) HAWKEN D R.
XX
XX Feder JN, Mintier G, Ramanathan CS, Hawken DR;
PI
XX WPI; 2004-603556/58.
XX
XX New isolated human G-coupled receptor (HGPRMY18), useful for diagnosing
PT or treating, e.g. disorders of the pituitary, hypogonadism, Cushing's
PT disease, osteoporosis, reproductive disorders, developmental disorders,
PT or colon cancer.
XX
XX Disclosure; SEQ ID NO 4; 76pp; English.
XX
XX This invention describes a novel human G-coupled receptor HGPRMY18
CC having the nucleic acid sequence of ATCC Accession No. PTA-2766. The
CC polynucleotide and polypeptide are useful for diagnosing or treating a
CC condition selected from endocrine-, gastrointestinal-, reproductive-,
CC pulmonary-, or neural-related disorders, e.g. disorders of the pituitary,
CC aberrant growth hormone synthesis and/or secretion, aberrant prolactin
CC synthesis and/or secretion, aberrant luteinizing hormone synthesis and/or
CC secretion, aberrant follicle-stimulating hormone synthesis and/or
CC secretion, aberrant thyroid-stimulating hormone synthesis and/or
CC secretion, aberrant adrenocorticotropin synthesis and/or secretion,
CC aberrant vasopressin secretion, aberrant oxytocin secretion, aberrant
CC growth, aberrant lactation, aberrant sexual characteristic development,
CC aberrant testosterone synthesis and/or secretion, aberrant oestrogen
CC synthesis and/or secretion, aberrant water homeostasis, hypogonadism,
CC Addison's disease, hypothyroidism, Cushing's disease, agromegaly,
CC gigantism, lethargy, osteoporosis, aberrant calcium homeostasis, aberrant
CC potassium homeostasis, reproductive disorders, developmental disorders,
CC colon cancer, related proliferative condition of the breast, prostate
CC cancer, related proliferative condition of the prostate, lung cancer,
CC related proliferative condition of the lung, renal tumours, or related
CC proliferative condition of the kidney. This sequence represents the human
CC orphan G-protein coupled receptor HGPRMY18 3'-UTR.
XX
XX Sequence 632 BP; 219 A; 113 C; 134 G; 166 T; 0 U; 0 Other;
SQ
XX
XX Query Match 0.8%; Score 37; DB 13; Length 632;
XX Best Local Similarity 100.0%; Pred. No. 0.0023;
XX Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
XX |||||||||||||||||||||||||||||||||||
Db 582 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 618
XX
XX RESULT 59
XX ACN04606/c
XX ID ACA04606 standard; cDNA; 647 BP.
XX
XX ACA04606;
XX
XX 28-MAY-2003 (first entry)
XX
XX cDNA encoding human membrane associated protein fragment #54.
XX
XX Human; ss; gene; microarray; membrane-associated protein; neuropathology;
XX immunopathology; pancreatic disease; cancer; diabetes; hyperlipidemia;
XX pancreatic cholest; Alzheimer's disease; Huntington's disease; sarcoma;
XX fibrocystic disease; leukaemia; adenocarcinoma; AIDS; allergy; anaemia;
XX asthma; gout; dementia.
XX
XX Homo sapiens.
XX
```

```
PN US6492505-B1.
XX
XX 10-DEC-2002.
XX
XX 31-JAN-2000; 2000US-00495050.
XX
XX 01-FEB-1999; 99US-0118318P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Reddy R, Guegler KJ, Au-Young J;
PI
XX WPI; 2003-327324/31.
XX
XX Claim 1; Col 69-72; 147pp; English.
XX
XX The invention relates to a combination comprising several polynucleotide
XX sequences comprising a fragment of gene encoding membrane-associated
XX proteins, receptors or ion channels. The combination is useful as a
XX probe, for research and diagnostic applications, for monitoring the
XX expression of several expressed polynucleotides, in the diagnosis and
XX monitoring of treatment of pancreatic disease, cancer, immunopathology or
XX neuropathology, for investigating an individual's predisposition to the
XX above disease, in genetic or gene expression analysis of polynucleotide
XX sequences, to investigate cellular responses to infection or drug
XX treatment, as hybridisable array elements in a microarray, to purify a
XX subpopulation of mRNAs, cDNAs or genomic fragments in a sample, in
XX diagnostics, prognostics and treatment regimens, in drug discovery and
XX development, in toxicological and carcinogenicity studies, and in
XX forensics or pharmacogenomics, to monitor the progression of disease, to
XX monitor the efficacy of treatment, to diagnose the conditions of the
XX pancreas e.g. diabetes, pancreatic cholera, hyperlipidemia or
XX fibrocystic disease, to diagnose a cancer e.g. leukaemia, adenocarcinoma
XX or sarcoma, to diagnose immunopathologies e.g. AIDS, allergies, anaemia,
XX asthma or gout, to diagnose neuropathologies e.g. Alzheimer's disease,
XX dementia or Huntington's disease, to rapidly screen large numbers of
XX candidate drug molecules and as query sequences against GenBank,
XX Swissprot, BLOCKS and PRINTS databases. The combination is employed to
XX fine tune the treatment regimen and thus the expression patterns
XX associated with undesirable side effects are avoided. The present
XX sequence represents a cDNA encoding a fragment of gene encoding human
XX membrane-associated proteins, receptors or ion channels
XX
XX Sequence 647 BP; 168 A; 107 C; 110 G; 262 T; 0 U; 0 Other;
SQ
XX
XX Query Match 0.8%; Score 37; DB 8; Length 647;
XX Best Local Similarity 100.0%; Pred. No. 0.0022;
XX Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
XX |||||||||||||||||||||||||||||||||||
Db 83 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 47
XX
XX RESULT 60
XX ACN87141/c
XX ID ACN87141 standard; DNA; 674 BP.
XX
XX ACN87141;
XX
XX 02-DEC-2004 (first entry)
XX
XX Breast cancer related marker, seq id 8291.
XX
XX Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.
XX
XX Homo sapiens.
XX
```


PN US2003099974-A1.
XX
PD 29-MAY-2003.
XX
PF 18-JUL-2002; 2002US-00198846.
XX
PR 18-JUL-2001; 2001US-0306220P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2003-787014/74.
XX
PT Novel isolated polypeptide associated with breast cancer, useful for
PT detecting presence of polypeptide in sample, as a marker for breast
PT cancer.
XX
PS Disclosure; SEQ ID NO 8291; 36pp; English.
XX
CC The invention relates to an isolated polypeptide (I) associated with
CC breast cancer which is encoded by a nucleic acid molecule comprising a
CC nucleotide sequence (SI). Further disclosed is an antibody that binds to
CC the polypeptide of the invention. The activity of the polypeptide of the
CC invention may be described as cytostatic. The antibody is useful for
CC detecting the presence of (I) in a sample. Nucleic acid molecules of the
CC invention are useful in the detection of breast tumours. (I) is useful as
CC a marker for breast cancer and in breast cancer therapy. Sequences given
CC in records ACN78851-ACN92934 represent nucleic acid markers associated
CC with breast cancer. Note: The sequence listing does not form part of the
CC specification but may be obtained in electronic format from the USPTO web
CC site at seqdata.uspto.gov/sequence.html?DocID=2003099974
XX
SQ Sequence 674 BP; 158 A; 128 C; 124 G; 147 T; 0 U; 117 Other;
Query Match 0.8%; Score 37; DB 11; Length 674;
Best Local Similarity 100.0%; Pred. No. 0.0022;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 89 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 53
RESULT 61
ID AAH33250 standard; cDNA; 782 BP.
XX
AC AAH33250;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:306.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.

DR P-PSDB; AAG73819.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 2426-2427; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated PS, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 782 BP; 197 A; 227 C; 179 G; 177 T; 0 U; 2 Other;
Query Match 0.8%; Score 37; DB 4; Length 782;
Best Local Similarity 100.0%; Pred. No. 0.0022;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 729 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 765
RESULT 62
ID AAT48447 standard; cDNA; 1218 BP.
XX
AC AAT48447;
XX
DT 11-APR-1997 (first entry)
XX
DE Human TGF-beta-like cytokine pCL13 variant b1.
XX
KW pCL13; clone 13; transforming growth factor-beta; TGF-beta;
KW immunosuppressive; cell differentiation; antiproliferative;
KW wound healing; vulnery; ischaemic injury; inflammation; cancer;
KW autoimmune disease; fibrotic disease; diagnosis; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FH CDS 11..937
FT /*tag= a
XX
XX WO9700958-A1.
XX
PD 09-JAN-1997.
XX
PF 24-JUN-1996; 96WO-AU000386.
XX
PR 22-JUN-1995; 95AU-00003706.
PR 23-AUG-1995; 95AU-00004990.
PR 09-FEB-1996; 96AU-00007983.
XX
PA (SVIN-) ST VINCENT'S HOSPITAL SYDNEY LTD.
XX
PI Breit S, Bootcov M;
XX
DR WPI; 1997-087385/08.

```
DR P-PSDB; AAM10671.
XX
XX Isolated nucleic acid encoding pCL13, new TGF-beta family member - has
PT immunosuppressant, cell differentiation promoting and anti-proliferative
PT activities, for treatment of wounds, ischaemic injury, cancer,
XX inflammatory disease etc.
XX
XX Example 17; Fig 20A; 73pp; English.
XX
XX cDNA clones b2, h1, u2, f1, a1, b1, d2 and dd2 (AAT48442-49) are variants
CC of clone 13 (AAT48438), which codes for pCL13 (AAM10662), a novel TGF-
CC beta-like cytokine whose activities include immunosuppression, down-
CC regulation of leucocyte extravasation and motility, promotion of cell
CC differentiation and inhibition of proliferation. The clones were obtd.
CC from a foetal lung library using a portion of the coding region of clone
CC 13 (see also AAT48438) as probe. They show substantial variation in the
CC 5' untranslated region but only minor differences in the coding region.
CC Translated coding regions are given in AAM10666-73. While the variants
CC may be important in e.g. transcriptional regulation they are untranslated
CC and hence cannot affect bioactivity
XX
XX Sequence 1218 BP; 249 A; 381 C; 361 G; 227 T; 0 U; 0 Other;
SQ
Query Match 0.8%; Score 37; DB 2; Length 1218;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1174 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1210

RESULT 63
AAT48446
ID AAT48446 standard; cDNA; 1227 BP.
XX
XX AAT48446;
XX
XX 11-APR-1997 (first entry)
XX
XX Human TGF-beta-like cytokine pCL13 variant a1.
XX
XX pCL13; clone 13; transforming growth factor-beta; TGF-beta;
XX immunosuppressive; cell differentiation; anti-proliferative;
XX wound healing; vulnerrary; ischaemic injury; inflammation; cancer;
XX autoimmune disease; fibrotic disease; diagnosis; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 20..946
XX /*tag= a
XX
XX MO9700958-A1.
XX
XX 09-JAN-1997.
XX
XX 24-JUN-1996; 96MO-AU000386.
XX
XX 22-JUN-1995; 95AU-00003706.
XX 23-AUG-1995; 95AU-00004990.
XX 09-FEB-1996; 96AU-00007983.
XX
XX (SVIN-) ST VINCENT'S HOSPITAL SYDNEY LTD.
XX
XX Breit S, Bootcov M;
XX
XX WPI; 1997-087385/08.
XX P-PSDB; AAM10670.
XX
XX Isolated nucleic acid encoding pCL13, new TGF-beta family member - has
PT immunosuppressant, cell differentiation promoting and anti-proliferative
PT activities, for treatment of wounds, ischaemic injury, cancer,
XX
```

```
PT inflammatory disease etc.
XX
XX Example 17; Fig 20A; 73pp; English.
XX
XX cDNA clones b2, h1, u2, f1, a1, b1, d2 and dd2 (AAT48442-49) are variants
CC of clone 13 (AAT48438), which codes for pCL13 (AAM10662), a novel TGF-
CC beta-like cytokine whose activities include immunosuppression, down-
CC regulation of leucocyte extravasation and motility, promotion of cell
CC differentiation and inhibition of proliferation. The clones were obtd.
CC from a foetal lung library using a portion of the coding region of clone
CC 13 (see also AAT48438) as probe. They show substantial variation in the
CC 5' untranslated region but only minor differences in the coding region.
CC Translated coding regions are given in AAM10666-73. While the variants
CC may be important in e.g. transcriptional regulation they are untranslated
CC and hence cannot affect bioactivity
XX
XX Sequence 1227 BP; 251 A; 384 C; 364 G; 228 T; 0 U; 0 Other;
SQ
Query Match 0.8%; Score 37; DB 2; Length 1227;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1183 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1219

RESULT 64
AAT48449
ID AAT48449 standard; cDNA; 1228 BP.
XX
XX AAT48449;
XX
XX 11-APR-1997 (first entry)
XX
XX Human TGF-beta-like cytokine pCL13 variant dd2.
XX
XX pCL13; clone 13; transforming growth factor-beta; TGF-beta;
XX immunosuppressive; cell differentiation; anti-proliferative;
XX wound healing; vulnerrary; ischaemic injury; inflammation; cancer;
XX autoimmune disease; fibrotic disease; diagnosis; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 21..947
XX /*tag= a
XX
XX MO9700958-A1.
XX
XX 09-JAN-1997.
XX
XX 24-JUN-1996; 96MO-AU000386.
XX
XX 22-JUN-1995; 95AU-00003706.
XX 23-AUG-1995; 95AU-00004990.
XX 09-FEB-1996; 96AU-00007983.
XX
XX (SVIN-) ST VINCENT'S HOSPITAL SYDNEY LTD.
XX
XX Breit S, Bootcov M;
XX
XX WPI; 1997-087385/08.
XX P-PSDB; AAM10673.
XX
XX Isolated nucleic acid encoding pCL13, new TGF-beta family member - has
PT immunosuppressant, cell differentiation promoting and anti-proliferative
PT activities, for treatment of wounds, ischaemic injury, cancer,
XX inflammatory disease etc.
XX
XX Example 17; Fig 20A; 73pp; English.
XX
XX cDNA clones b2, h1, u2, f1, a1, b1, d2 and dd2 (AAT48442-49) are variants
CC
```

CC of clone 13 (AA1748438), which codes for pCL13 (AAW10662), a novel TGF-
CC beta-like cytokine whose activities include immunosuppression, down-
CC regulation of leucocyte extravasation and motility, promotion of cell
CC differentiation and inhibition of proliferation. The clones were obd.
CC from a foetal lung library using a portion of the coding region of clone
CC 13 (see also AA1748438) as probe. They show substantial variation in the
CC 5' untranslated region but only minor differences in the coding region.
CC Translated coding regions are given in AAW10666-73. While the variants
CC may be important in e.g. transcriptional regulation they are untranslated
CC and hence cannot affect bioactivity
XX
SQ Sequence 1228 BP; 251 A; 386 C; 363 G; 228 T; 0 U; 0 Other;
Query Match 0.8%; Score 37; DB 2; Length 1228;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1184 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1220
RESULT 65
AAD51362
ID AAD51362 standard; cDNA; 1345 BP.
XX
AC AAD51362;
XX
DT 16-APR-2003 (first entry)
XX
DE Soybean flavonol synthase cDNA.
XX
KM Isoflavonoid; transgenic; C1 myb; R myc-type transcription factor; feed;
KM food; beverage; soybean; flavonol synthase; enzyme; gene; ss.
XX
OS Glycine max.
XX
PN WO2002101023-A2.
XX
PD 19-DEC-2002.
XX
PF 13-JUN-2002; 2002WO-US021107.
XX
PR 13-JUN-2001; 2001US-0297981P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Odel1 JT, Yu X;
XX
DR WPI; 2003-148796/14.
XX
PT Altering isoflavonoid profile of isoflavonoid-producing plant, by using
PT recombinant expression constructs having promoters linked to nucleic acid
PT fragments encoding C1 myb and R myc-type transcription factors.
XX
PS Example 6; Col 74; 38pp; English.
XX
CC The invention relates to a method for altering isoflavonoid profile of
CC isoflavonoid-producing plant, by using recombinant expression constructs
CC having promoters linked to nucleic acid fragments encoding C1 myb and R
CC myc-type transcription factors. The invention is useful for altering
CC isoflavonoid profile of isoflavonoid-producing plant such as soybean,
CC clover, mung bean, lentil, hairy vetch, alfalfa, lupine, sugar beet, and
CC snow pea. It is useful for producing an isoflavonoid-containing product
CC by cracking the seeds to remove the meats from the hulls, and flaking the
CC meats to obtain the desired flake thickness. Isoflavonoid-containing
CC product is useful in feeds, foods and/or beverages. It is useful for
CC producing pills, tablets or capsules. The present sequence is soybean
CC flavonol synthase cDNA used to illustrate the method of the invention
CC
SQ Sequence 1345 BP; 457 A; 228 C; 312 G; 348 T; 0 U; 0 Other;
Query Match 0.8%; Score 37; DB 8; Length 1345;

Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1296 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1332
RESULT 66
AA1748444
ID AA1748444 standard; cDNA; 1441 BP.
XX
AC AA1748444;
XX
DT 11-APR-1997 (first entry)
XX
DE Human TGF-beta-like cytokine pCL13 variant u2.
XX
XX pCL13; clone 13; transforming growth factor-beta; TGF-beta;
XX immunosuppressive; cell differentiation; antiproliferative;
XX wound healing; vulnary; ischaemic injury; inflammation; cancer;
XX autoimmune disease; fibrotic disease; diagnosis; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH 234..1160
FT CDS /*tag= a
FT
XX
XX WO9700958-A1.
XX
XX 09-JAN-1997.
XX
XX 24-JUN-1996; 96WO-AU000386.
XX
XX 22-JUN-1995; 95AU-00003706.
XX 23-AUG-1995; 95AU-00004990.
XX 09-FEB-1996; 96AU-00007983.
XX
PA (SVIN-) ST VINCENT'S HOSPITAL SYDNEY LTD.
XX
XX Breit S, Bootcov M;
XX
XX WPI; 1997-087385/08.
XX
XX P-PSDB; AAW10668.
XX
XX Isolated nucleic acid encoding pCL13, new TGF-beta family member - has
XX immunosuppressant, cell differentiation promoting and anti:proliferative
XX activities, for treatment of wounds, ischaemic injury, cancer,
XX inflammatory disease etc.
XX
XX Example 17; Fig 20A; 73pp; English.
XX
XX cDNA clones b2, h1, u2, f1, a1, b1, d2 and dd2 (AA1748442-49) are variants
XX of clone 13 (AA1748438), which codes for pCL13 (AAW10662), a novel TGF-
XX beta-like cytokine whose activities include immunosuppression, down-
XX regulation of leucocyte extravasation and motility, promotion of cell
XX differentiation and inhibition of proliferation. The clones were obd.
XX from a foetal lung library using a portion of the coding region of clone
XX 13 (see also AA1748438) as probe. They show substantial variation in the
XX 5' untranslated region but only minor differences in the coding region.
XX Translated coding regions are given in AAW10666-73. While the variants
XX may be important in e.g. transcriptional regulation they are untranslated
XX and hence cannot affect bioactivity
XX
SQ Sequence 1441 BP; 300 A; 451 C; 428 G; 262 T; 0 U; 0 Other;
Query Match 0.8%; Score 37; DB 2; Length 1441;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
|||||

RESULT 69
ADA98154
ID ADA98154 standard; cDNA; 1663 BP.
XX
XX
AC ADA98154;
XX
XX 20-NOV-2003 (first entry)
XX
DE Human secreted protein cDNA sequence #248.
XX
XX human; secreted protein; cardiovascular disorder; arrhythmia;
KM atherosclerosis; stroke; endocarditis; congestive heart failure;
KM rheumatic heart disease; cardiomyopathy; haemorrhoids; varicose veins;
KM migraine; thrombosis; neural disorder; immune system disorder;
KM muscular disorder; reproductive disorder; gastrointestinal disorder;
KM pulmonary disorder; renal disorder; proliferative disorder; cancer; gene;
KM SB.
XX
XX Homo sapiens.
OS
XX
XX WO2003004623-A2.
XX
XX 16-JAN-2003.
XX
XX 26-MAR-2002; 2002WO-US009922.
XX
XX 27-MAR-2001; 2001US-0278650P.
XX 12-SEP-2001; 2001US-00950082.
XX 12-SEP-2001; 2001US-00950083.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
PI
XX WPI; 2003-247946/24.
XX
XX New human secreted polypeptide and nucleic acid molecules, useful for
PT diagnosing, preventing, prognosticating or treating cardiovascular
PT disorders (e.g. arrhythmia, atherosclerosis, cardiomyopathy, or
PT thrombosis).
XX
XX Claim 1; SEQ ID NO 258; 1572pp; English.
XX
XX The invention comprises the amino acid and coding sequence of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful in the treatment of cardiovascular disorders, such as: arrhythmia,
CC atherosclerosis, stroke, endocarditis, congestive heart failure,
CC rheumatic heart disease, cardiomyopathy, haemorrhoids, varicose veins,
CC migraine, or thrombosis. The DNA and protein sequences may also be used
CC for treating or preventing: neural disorders, immune system disorders,
CC muscular disorders, reproductive disorders, gastrointestinal disorders,
CC pulmonary disorders, renal disorders, proliferative disorders and/or
CC cancerous diseases. The present cDNA sequence encodes a human secreted
CC protein of the invention. NOTE: The present sequence is shown on the WIPO
CC website.
XX
XX Sequence 1663 BP; 549 A; 250 C; 313 G; 551 T; 0 U; 0 Other;
XX
XX
Query Match 0.8%; Score 37; DB 8; Length 1663;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AC ABT16860;
XX
XX
XX 03-APR-2003 (first entry)
XX
XX Human secreted protein gene sequence - SEQ ID NO 109.
DE
XX
XX Human; gene; ds; protein therapy; immediate hypersensitivity disease;
KM allergic disorder; asthmatic disorder; gene therapy; secreted protein;
KM hay fever; allergic conjunctivitis; allergic rhinitis;
KM binding partner identification; chromosome identification;
KM radiation hybrid mapping; long-range restriction mapping.
XX
XX Homo sapiens.
OS
XX
XX WO200277188-A2.
XX
XX 03-OCT-2002.
XX
XX 26-MAR-2002; 2002WO-US009239.
XX
XX 27-MAR-2001; 2001US-0278650P.
XX 12-SEP-2001; 2001US-00950082.
XX 12-SEP-2001; 2001US-00950083.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
PI
XX WPI; 2003-175010/17.
XX
XX Use of human secreted proteins and nucleic acids for preparing a
PT diagnostic or pharmaceutical composition for diagnosing or treating
PT allergic or asthmatic disorders, e.g. asthma, hay fever, or allergic
PT conjunctivitis or rhinitis.
XX
XX Claim 7; Page 620; 823pp; English.
XX
XX The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for the diagnosis and treatment of allergic disorders, asthmatic
CC disorders and immediate hypersensitivity diseases (e.g. hay fever,
CC allergic conjunctivitis and allergic rhinitis). The proteins of the
CC invention are also useful for identifying a binding partner. The nucleic
CC acids of the invention are also useful for chromosome identification.
CC radiation hybrid mapping or long-range restriction mapping. The present
CC DNA sequence encodes a human secreted protein of the invention
XX
XX Sequence 1663 BP; 549 A; 250 C; 313 G; 551 T; 0 U; 0 Other;
XX
XX
Query Match 0.8%; Score 37; DB 10; Length 1663;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
XX
XX 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 1595 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1631

RESULT 71
AB267274
ID AB267274 standard; cDNA; 1663 BP.
XX
XX
AC AB267274;
XX
XX 26-MAR-2003 (first entry)
XX
XX Human secreted protein encoding cDNA SEQ ID NO 394.
DE
XX
XX Human; secreted protein; noctropic; neuroprotective; cyrostatic;
KM virucide; dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
KM vulnerary; antibacterial; antiparkinsonian; antischist; antianemic;
KM antiarthritic; cancer; antineumatic; hepatotropic; cerebroprotective;
KM antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;
KW vulnary; anticonvulsant; cancer; immune disorder; cardiovascular disorder;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
OS Homo sapiens.
XX
XX WO20061624-A1.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US008980.
XX
XX 09-APR-1999; 99US-0128700P.
XX
XX 20-JAN-2000; 2000US-0176930P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-656324/63.
XX
XX P-PSDB; AAB52145.
XX
XX New nucleic acid molecules encoding human secreted proteins, used in
XX preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
XX Parkinson's diseases and cancers.
XX
XX Claim 1; Page 420-421, 478pp; English.
XX
XX The invention relates to the isolation of genes AAC96900-C96947 encoding
XX 48 human secreted proteins AAB52104-B52150. The genes can be used to
XX generate fusion proteins by linking to the gene for the human
XX immunoglobulin G Fc portion (SEQID) for increasing the stability of the
XX fusion protein as compared to the human protein only. The genes and
XX proteins are useful for preventing, ameliorating or treating medical
XX conditions, e.g. by protein or gene therapy. The genes are isolated from
XX a range of human tissues disclosed in the specification. The nucleic
XX acids, proteins, antibodies and (ant)agonists are useful in the
XX diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer, and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)
XX immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
XX cardiovascular disorders such as myocardial ischaemias; (d) wound healing
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections
SQ Sequence 1703 BP; 442 A; 386 C; 382 G; 491 T; 0 U; 2 Other;
Query Match 0.8%; Score 37; DB 3; Length 1703;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4336 CTGTTAAAAA
Db 1643 CTGTTAAAAA
RESULT 74
ABL90376
ID ABL90376 standard; cDNA; 1703 BP.
XX
XX ABL90376;
XX
XX 24-MAY-2002 (first entry)
XX
XX Human polynucleotide SEQ ID NO 938.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;
KW

KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.
OS Homo sapiens.
XX
XX WO200190304-A2.
XX
XX 29-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-US016450.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
XX
XX WPI; 2002-122018/16.
XX
XX P-PSDB; AAB89967.
XX
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
XX prevention of neural, immune system, muscular, reproductive, and
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative
XX disorders.
XX
XX Claim 4; SEQ ID NO 938; 2081pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
XX (AAB89040-AAB90444) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow,
XX breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
XX cardiovascular disorders such as myocardial ischaemias; (d) wound healing
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 1703 BP; 442 A; 386 C; 381 G; 491 T; 0 U; 3 Other;
Query Match 0.8%; Score 37; DB 6; Length 1703;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4336 CTGTTAAAAA
Db 1643 CTGTTAAAAA
RESULT 75
AAL53985
ID AAL53985 standard; cDNA; 1706 BP.
XX
XX AAL53985;
XX
XX 18-FEB-2003 (first entry)
XX
XX cDNA encoding a soybean high affinity ammonium transporter protein.
XX
XX Herbicide; ammonium transporter protein; herbicide; transgenic plant;
XX soybean; gene; ss.
XX
XX Glycine max.
XX
XX Key Location/Qualifiers
FH

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2005, 10:44:45 ; Search time 2729 Seconds
(without alignments)
9547.160 Million cell updates/sec

Title: US-10-054-935-1

Perfect score: 4372
Sequence: 1 cagtcctcgaccccccgcac.....aaaaaaaaaaaaaaaa 4372

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
5552208 seqs, 297965951 residues

Word size : 15
Total number of hits satisfying chosen parameters: 471715

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubppa/US05_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubppa/US04_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubppa/US03_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubppa/US02_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubppa/US01_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubppa/US00_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubppa/US05_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubppa/US04_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubppa/US03_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubppa/US02_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubppa/US01_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubppa/US00_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubppa/US11_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.seq:*
- 22: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4372	100.0	4372	US-10-054-935-1	Sequence 1, Appl1
2	4372	100.0	4372	US-10-144-194A-65	Sequence 65, Appl1
3	4372	100.0	4372	US-10-491-566-65	Sequence 65, Appl1
4	1260	28.8	1509	US-09-397-945-79	Sequence 79, Appl1
5	1260	28.8	1509	US-10-653-595-79	Sequence 79, Appl1
6	903	20.7	1545	US-10-641-643-167	Sequence 167, Appl1
7	837	19.1	1529	US-09-925-302-212	Sequence 212, Appl1
8	837	19.1	1529	US-09-925-302-212	Sequence 212, Appl1
9	487	11.1	1290	US-10-641-643-322	Sequence 322, Appl1
10	436	10.0	600	US-10-242-535A-51339	Sequence 51339, A
11	436	10.0	600	US-10-085-783A-51339	Sequence 51339, A

12	410	9.4	451	17	US-10-242-535A-58053	Sequence 58053, A
13	410	9.4	451	17	US-10-085-783A-58053	Sequence 58053, A
14	357	8.2	365	10	US-09-918-995-30064	Sequence 30064, A
15	281	6.4	400	17	US-10-242-535A-8833	Sequence 8833, Ap
16	281	6.4	400	17	US-10-085-783A-8833	Sequence 8833, Ap
17	260	5.9	273	18	US-10-425-115-150265	Sequence 150265, A
18	239	5.5	245	17	US-10-242-535A-43646	Sequence 43646, A
19	239	5.5	245	17	US-10-085-783A-43646	Sequence 43646, A
20	230	5.3	220	9	US-09-998-598-2127	Sequence 2127, Ap
21	214	4.9	440	17	US-10-242-535A-31928	Sequence 31928, A
22	214	4.9	440	17	US-10-085-783A-31928	Sequence 31928, A
23	213	4.9	411	10	US-09-918-995-34149	Sequence 34149, A
24	152	3.5	542	9	US-09-736-457-1554	Sequence 1554, Ap
25	152	3.5	542	9	US-09-902-941-1554	Sequence 1554, Ap
26	152	3.5	542	9	US-09-849-626-1554	Sequence 1554, Ap
27	152	3.5	542	14	US-10-017-754-1554	Sequence 1554, Ap
28	152	3.5	542	16	US-10-113-872-1554	Sequence 1554, Ap
29	152	3.5	542	17	US-10-283-017-1554	Sequence 1554, Ap
30	134	3.1	211	18	US-10-357-930-56465	Sequence 56465, A
31	110	2.5	615	9	US-09-917-800A-725	Sequence 725, Ap
32	63	1.4	346	9	US-09-983-965-2328	Sequence 2328, Ap
33	60	1.4	60	10	US-09-908-975-22320	Sequence 22320, A
34	51	1.2	580	9	US-09-917-800A-16	Sequence 16, Appl1
35	51	1.2	530	17	US-10-191-803-451	Sequence 451, Appl1
36	49	1.1	156	15	US-10-106-698-3799	Sequence 3799, Ap
37	40	0.9	453	18	US-10-425-115-167742	Sequence 167742, A
38	40	0.9	642	18	US-10-425-115-95727	Sequence 95727, A
39	40	0.9	2637	17	US-10-424-599-109018	Sequence 109018, A
40	39	0.9	2637	17	US-10-424-599-109018	Sequence 109018, A
41	39	0.9	2637	17	US-10-437-963-13051	Sequence 13051, A
42	39	0.9	506	18	US-10-437-963-62466	Sequence 62466, A
43	38	0.9	2472	18	US-10-437-963-19880	Sequence 19880, A
44	38	0.9	376	18	US-10-425-115-43467	Sequence 43467, A
45	38	0.9	827	9	US-09-982-809-20	Sequence 20, Appl1
46	38	0.9	1654	9	US-09-925-302-156	Sequence 156, Appl1
47	38	0.9	1955	18	US-10-425-115-14882	Sequence 14882, A
48	38	0.9	2200	18	US-10-425-115-152967	Sequence 152967, A
49	38	0.9	2840	18	US-10-437-963-66987	Sequence 66987, A
50	37	0.8	241	9	US-09-960-352-12974	Sequence 12974, A
51	37	0.8	243	18	US-10-357-930-56658	Sequence 56658, A
52	37	0.8	2647	18	US-10-425-115-4861	Sequence 4861, Ap
53	37	0.8	333	18	US-10-425-115-108211	Sequence 108211, A
54	37	0.8	351	18	US-10-425-115-60416	Sequence 60416, A
55	37	0.8	374	18	US-10-425-115-101254	Sequence 101254, A
56	37	0.8	394	17	US-10-424-599-112320	Sequence 112320, A
57	37	0.8	410	10	US-09-918-995-8826	Sequence 8826, Ap
58	37	0.8	418	10	US-09-918-995-8826	Sequence 8826, Ap
59	37	0.8	421	18	US-10-425-115-142294	Sequence 142294, A
60	37	0.8	444	17	US-10-424-599-3745	Sequence 3745, Ap
61	37	0.8	450	10	US-09-918-995-10500	Sequence 10500, A
62	37	0.8	464	18	US-10-425-115-51597	Sequence 51597, A
63	37	0.8	553	18	US-10-425-115-50595	Sequence 50595, A
64	37	0.8	632	10	US-09-992-331-4	Sequence 4, Appl1
65	37	0.8	632	15	US-10-262-313-4	Sequence 4, Appl1
66	37	0.8	632	17	US-10-768-878-4	Sequence 4, Appl1
67	37	0.8	642	18	US-10-424-599-66141	Sequence 66141, A
68	37	0.8	643	18	US-10-425-115-1661	Sequence 1661, Ap
69	37	0.8	647	15	US-10-313-542-54	Sequence 54, Appl1
70	37	0.8	648	18	US-10-425-115-15766	Sequence 15766, A
71	37	0.8	661	18	US-10-437-963-37096	Sequence 37096, A
72	37	0.8	674	14	US-10-198-846-8291	Sequence 8291, Ap
73	37	0.8	698	18	US-10-437-963-32884	Sequence 32884, A
74	37	0.8	722	18	US-10-425-115-164365	Sequence 164365, A
75	37	0.8	762	15	US-10-106-698-316	Sequence 316, Appl1
76	37	0.8	1128	18	US-10-425-115-136494	Sequence 136494, A
77	37	0.8	1256	18	US-10-425-115-11638	Sequence 11638, A
78	37	0.8	1337	18	US-10-425-115-157597	Sequence 157597, A
79	37	0.8	1335	16	US-10-171-174A-13	Sequence 13, Appl1
80	37	0.8	1354	18	US-10-425-115-177452	Sequence 177452, A
81	37	0.8	1703	17	US-10-264-237-938	Sequence 938, Appl1
82	37	0.8	1706	13	US-10-033-109-3	Sequence 3, Appl1
83	37	0.8	1722	15	US-10-177-390-17	Sequence 17, Appl1
84	37	0.8	1775	18	US-10-437-963-41567	Sequence 41567, A

```
85 37 0.8 1809 17 US-10-424-599-75832 Sequence 75832, A
86 37 0.8 1883 10 US-09-809-391-170 Sequence 170, App
87 37 0.8 1883 10 US-09-882-171-170 Sequence 170, App
88 37 0.8 1883 17 US-10-164-861-170 Sequence 170, App
89 37 0.8 2030 17 US-10-321-802-27 Sequence 27, App1
90 37 0.8 2237 18 US-10-437-963-36575 Sequence 36575, A
91 37 0.8 2307 18 US-10-437-963-33485 Sequence 33485, A
92 37 0.8 2419 18 US-10-425-115-83962 Sequence 83962, A
93 37 0.8 2439 18 US-10-357-930-21114 Sequence 21114, A
94 37 0.8 2522 18 US-10-357-930-21114 Sequence 21114, A
95 37 0.8 2532 18 US-10-357-930-23487 Sequence 23487, A
96 37 0.8 2532 18 US-10-357-930-26954 Sequence 26954, A
97 37 0.8 2522 18 US-10-357-930-29373 Sequence 29373, A
98 37 0.8 2807 18 US-10-755-889-519 Sequence 519, App
99 37 0.8 2808 17 US-10-295-027-1001 Sequence 1001, App
100 37 0.8 2915 10 US-09-374-046A-101 Sequence 101, App
```

ALIGNMENTS

```
RESULT 1
US-10-054-935-1
; Sequence 1, Application US/10054935
; Publication No. US20030143546A1
;
GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: BREAST CANCER TRANSCRIPTION FACTOR GENE AND USES
; FILE REFERENCE: 16U 107 R1
; CURRENT APPLICATION NUMBER: US/10/054,935
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 1
; LENGTH: 4372
; TYPE: DNA
; ORGANISM: Homo sapiens
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(1922)
; OTHER INFORMATION:
US-10-054-935-1
```

```
Query Match 100.0%; Score 4372; DB 15; Length 4372;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CAGTCTCGACCCCGCAGCTGCGCCCTCCCAACCCCTCCTCGCGCTCGATGCCCG 60
Db 1 CAGTCTCGACCCCGCAGCTGCGCCCTCCCAACCCCTCCTCGCGCTCGATGCCCG 60
QY 61 CGGTGCTCGGACCACTATGACATGATCGCGGTTCGAAGCGCGCGCCCTG 120
Db 61 CGGTGCTCGGACCACTATGACATGATCGCGGTTCGAAGCGCGCGCCCTG 120
QY 121 CGGCGCGCAATCTTGAGCAGCTGAGCTAGCAGCGGCGTGGCGCTGGCGCCG 180
Db 121 CGGCGCGCAATCTTGAGCAGCTGAGCTAGCAGCGGCGTGGCGCTGGCGCCG 180
QY 181 AGGACGAGCTTGAGGCGCGGAGCCCACTTCTCCCGCGCAACCTTAACCTCAAGAGC 240
Db 181 AGGACGAGCTTGAGGCGCGGAGCCCACTTCTCCCGCGCAACCTTAACCTCAAGAGC 240
QY 241 CGGCGCGCGCTGCGCTCTCTCCAGGCGGAGCCCGCGCTTCCCGCGCGCTGCG 300
Db 241 CGGCGCGCGCTGCGCTCTCTCCAGGCGGAGCCCGCGCTTCCCGCGCGCTGCG 300
QY 301 GCGGCAAGGCGCGGCGCTTGTACTCTCGCGCGGCGGCGCCCGCGGCAAGAGAGA 360
Db 301 GCGGCAAGGCGCGGCGCTTGTACTCTCGCGCGGCGGCGCCCGCGGCAAGAGAGA 360
QY 361 GCTGGGCGGTTGGGTCCCTTGGCTTCCGCTCCCGCGGCAACGAGCGGCACTTG 420
Db 361 GCTGGGCGGTTGGGTCCCTTGGCTTCCGCTCCCGCGGCAACGAGCGGCACTTG 420
```

```
Db 361 GCTGGGCGGTTGGGTCCCTTGGCTTCCGCTCCCGCGGCAACGAGCGGCACTTG 420
QY 421 GGGGGAGGCTTGGCGGAGCGGAGCGGCTGCAAGCCCCCGGCCCAATATCAGCGGTG 480
Db 421 GGGGGAGGCTTGGCGGAGCGGAGCGGCTGCAAGCCCCCGGCCCAATATCAGCGGTG 480
QY 481 TGGCCATTGAGACGGGCTCTCTGAGGCGGCGCAAGAGCTTACGCTGGGCTGGGG 540
Db 481 TGGCCATTGAGACGGGCTCTCTGAGGCGGCGCAAGAGCTTACGCTGGGCTGGGG 540
QY 541 ACNAGGTTGGGCGGCGCTCCCGCTGACCGGCTCGAGCCCGGCGGAGCCCGCAC 600
Db 541 ACNAGGTTGGGCGGCGCTCCCGCTGACCGGCTCGAGCCCGGCGGAGCCCGCAC 600
QY 601 TACCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 660
Db 601 TACCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 660
QY 661 GCGAGGCGGATGGAAGATATGAGGAAGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 661 GCGAGGCGGATGGAAGATATGAGGAAGGCGGCGGCGGCGGCGGCGGCGGCGG 720
QY 721 CCTCAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db 721 CCTCAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 780
QY 781 AGCAGCAGCAGCAGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 840
Db 781 AGCAGCAGCAGCAGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 840
QY 841 ACACGCTCTGCTCGGATTTGAACGATGGAAGGCGGCGGCGGCGGCGGCGGCG 900
Db 841 ACACGCTCTGCTCGGATTTGAACGATGGAAGGCGGCGGCGGCGGCGGCGGCG 900
QY 901 ACAGAGAGAGAGAGCAGAGCTGTTCAAGGCTATGAATGAAGAGAGAGAGAGAG 960
Db 901 ACAGAGAGAGAGAGCAGAGCTGTTCAAGGCTATGAATGAAGAGAGAGAGAGAG 960
QY 961 AGCTATCTGAGAAATTAATGAGAGTGCAGCGGAGCTTTCCGAGATCCGAGATC 1020
Db 961 AGCTATCTGAGAAATTAATGAGAGTGCAGCGGAGCTTTCCGAGATCCGAGATC 1020
QY 1021 TGGCTCCCAAGCCCTTCTCATGTGGCGGAGTGGAGAGGAGCAATTAAGAAATCC 1080
Db 1021 TGGCTCCCAAGCCCTTCTCATGTGGCGGAGTGGAGAGGAGCAATTAAGAAATCC 1080
QY 1081 TTGGAAGTACAGAAAGAGAGCTCTGTTAAAGCTGCTCTGAAATTTTCAAAAGTCA 1140
Db 1081 TTGGAAGTACAGAAAGAGAGCTCTGTTAAAGCTGCTCTGAAATTTTCAAAAGTCA 1140
QY 1141 AAACAAAACTCTTAAGCACTCTCTTAATTAAGAGAACCTGTGTGTTCTTAATCGAAA 1200
Db 1141 AAACAAAACTCTTAAGCACTCTCTTAATTAAGAGAACCTGTGTGTTCTTAATCGAAA 1200
QY 1201 CTGTTTGTAACTGTAAATGAGAGCGCAAGAAACCCCAAGAAAGCCCGGTCTTCACTG 1260
Db 1201 CTGTTTGTAACTGTAAATGAGAGCGCAAGAAACCCCAAGAAAGCCCGGTCTTCACTG 1260
QY 1261 ACACCCCAAGAGCTCTCACTCCCAAAAGGAGCCAGACCCATCCCAAGAGAGAG 1320
Db 1261 ACACCCCAAGAGCTCTCACTCCCAAAAGGAGCCAGACCCATCCCAAGAGAGAG 1320
QY 1321 CTTTCTCAAGTAGATAGAAATTTGCGTACTTCTTCAACACAGAAATGTATTTGTGTC 1380
Db 1321 CTTTCTCAAGTAGATAGAAATTTGCGTACTTCTTCAACACAGAAATGTATTTGTGTC 1380
QY 1381 GTTGGACAGAGCTCCCGCATCACCGTTACATTAAGGAAATCTTCCAAAGAGAGAG 1440
Db 1381 GTTGGACAGAGCTCCCGCATCACCGTTACATTAAGGAAATCTTCCAAAGAGAGAG 1440
QY 1441 AGACTGTACAGAGTGTGATGTCATCAAGTGTTCAGAGAGAACTTCACTTGTGCTG 1500
Db 1441 AGACTGTACAGAGTGTGATGTCATCAAGTGTTCAGAGAGAACTTCACTTGTGCTG 1500
```


D	1081	TTGAAATACAGAAAGAAAGAACTCTGTTAAAGCGGCTCCGAAATTTTCAAAAGTCA	1140
Q	1141	AAACAAAAATCTCTAGACACTCTCTATTTAAAGAAACCTGNGTTCCTTATCTGAAA	1200
D	1141	AAACAAAAATCTCTAGACACTCTCTATTTAAAGAAACCTGNGTTCCTTATCTGAAA	1200
Q	1201	CTGTTTGAACGTGAATTGAGAGCCAAAGAAACCCGAGAAAGCCCGGTCTTCACTGG	1260
D	1201	CTGTTTGAACGTGAATTGAGAGCCAAAGAAACCCGAGAAAGCCCGGTCTTCACTGG	1260
Q	1261	ACACCCCAACCAAGCTCTCCATCTCCCAAAAGGAGCCAGACCCATCTCCAGAGAAAG	1320
D	1261	ACACCCCAACCAAGCTCTCCATCTCCCAAAAGGAGCCAGACCCATCTCCAGAGAAAG	1320
Q	1321	CCCTTCAAGTGAATGAGATTTGCGGTAACCTTTCACACAGAAATGATTTGTGTC	1380
D	1321	CCCTTCAAGTGAATGAGATTTGCGGTAACCTTTCACACAGAAATGATTTGTGTC	1380
Q	1381	GTTGACACAGCTCCCCCATCAACCTTACATTAACGGGAATCCTCTCAAGAGAGAGG	1440
D	1381	GTTGACACAGCTCCCCCATCAACCTTACATTAACGGGAATCCTCTCAAGAGAGAGG	1440
Q	1441	AGA CTGAGCAAGGTGTCTGATGCAATCAAGTGTGAGAGAAATTCAGTCTTGCTG	1500
D	1441	AGA CTGAGCAAGGTGTCTGATGCAATCAAGTGTGAGAGAAATTCAGTCTTGCTG	1500
Q	1501	TTCTCTTCTGAGAGGACCACTCACTGATGAGCTCTTAAGGACCCCAATCTTCAAGCTTT	1560
D	1501	TTCTCTTCTGAGAGGACCACTCACTGATGAGCTCTTAAGGACCCCAATCTTCAAGCTTT	1560
Q	1561	TGAGAAACTGATGACAGTGTGTTTCGAAGCGGCATGCAAACTGAGCTGATGAGA	1620
D	1561	TGAGAAACTGATGACAGTGTGTTTCGAAGCGGCATGCAAACTGAGCTGATGAGA	1620
Q	1621	AGAGAAAGAAAGATGGGATATTCAGAGATCAGGAGAACAAAGATTTTACAGGACTGC	1680
D	1621	AGAGAAAGAAAGATGGGATATTCAGAGATCAGGAGAACAAAGATTTTACAGGACTGC	1680
Q	1681	AGCTCAGAAATGTATTAATAAGAAAGAAATTCAGAAATCTGAGCTGAGATTAACCTCATTTT	1740
D	1681	AGCTCAGAAATGTATTAATAAGAAAGAAATTCAGAAATCTGAGCTGAGATTAACCTCATTTT	1740
Q	1741	TCCCTGAGCAGATGATGTAAGAAAGTTGATGATTAACCCCTTCTGCTGTTGATGAT	1800
D	1741	TCCCTGAGCAGATGATGTAAGAAAGTTGATGATTAACCCCTTCTGCTGTTGATGAT	1800
Q	1801	TTGAGCAGCAATTAACCAAAATTAATCCACAGAAATTTGAGCTACCTGCTGATGAGC	1860
D	1801	TTGAGCAGCAATTAACCAAAATTAATCCACAGAAATTTGAGCTACCTGCTGATGAGC	1860
Q	1861	GTAAGCAGATGATGTAAGATTCAGAAAGCAACCTCAACCGAGCTGTGAGAAAT	1920
D	1861	GTAAGCAGATGATGTAAGATTCAGAAAGCAACCTCAACCGAGCTGTGAGAAAT	1920
Q	1921	AGCTGAGTGGCAAGAACCCGCTTCTGATGATGTAAGATTCAGATTTCCCGAGAGTGC	1980
D	1921	AGCTGAGTGGCAAGAACCCGCTTCTGATGATGTAAGATTCAGATTTCCCGAGAGTGC	1980
Q	1981	AGAGACCTGATATGTAAGCTTGTCTCTACATATGTTATCACTCGCTGATTAACCTT	2040
D	1981	AGAGACCTGATATGTAAGCTTGTCTCTACATATGTTATCACTCGCTGATTAACCTT	2040
Q	2041	TCATATCTTCTGATCTTGTGTTTCAATTAATCTGATTTTCAAAATCTTTCATTCGGCT	2100
D	2041	TCATATCTTCTGATCTTGTGTTTCAATTAATCTGATTTTCAAAATCTTTCATTCGGCT	2100
Q	2101	AATGATGATTAAGAGGATGATGGATTTCTTTTCTTTTGGAAATGGGCTCTC	2160
D	2101	AATGATGATTAAGAGGATGATGGATTTCTTTTCTTTTGGAAATGGGCTCTC	2160
Q	2161	AAGTAAAGCTATAGATGGCAGATTCAGAAATTCAGAGGCTGTGTTCTATACATTTGC	2220
D	2161	AAGTAAAGCTATAGATGGCAGATTCAGAAATTCAGAGGCTGTGTTCTATACATTTGC	2220
D	2161	AAGTAAAGCTATAGATGGCAGATTCAGAAATTCAGAGGCTGTGTTCTATACATTTGC	2220
Q	2221	CTATGTTAAAGGGGTAAAGGGCTCTTTCATTAAGACATGAGAGATGAGAGAGCCCT	2280
D	2221	CTATGTTAAAGGGGTAAAGGGCTCTTTCATTAAGACATGAGAGATGAGAGAGCCCT	2280
Q	2281	TCTTTAGAGCTGTGCTGATGAGCACTTCTCACCTGTGTAACCTCTTATAGTGG	2340
D	2281	TCTTTAGAGCTGTGCTGATGAGCACTTCTCACCTGTGTAACCTCTTATAGTGG	2340
Q	2341	GTAATGATTTTAAACCTTAATAAACAACCTCAACATGAGCTTTAGAGACAG	2400
D	2341	GTAATGATTTTAAACCTTAATAAACAACCTCAACATGAGCTTTAGAGACAG	2400
Q	2401	AAGAGATGACAAAGTGAAGCGATGAGAGCAAGCATTTCAAGAGTGAAGAAAGACATCG	2460
D	2401	AAGAGATGACAAAGTGAAGCGATGAGAGCAAGCATTTCAAGAGTGAAGAAAGACATCG	2460
Q	2461	GAGAGTGTGATGATTAATGTTGTAAGAAAGATGTTTCAATTTGAACTATCTGTGATAC	2520
D	2461	GAGAGTGTGATGATTAATGTTGTAAGAAAGATGTTTCAATTTGAACTATCTGTGATAC	2520
Q	2521	AGTATGTGGAAGGATGTTTGTGCTGTGATTAATTTTCAATTTGATGATTAACATTTTC	2580
D	2521	AGTATGTGGAAGGATGTTTGTGCTGTGATTAATTTTCAATTTGATGATTAACATTTTC	2580
Q	2581	TTTACTGCTCAAAACCAAAATCTTGGAAAGAAAGTGGGATGCTTATTTCAAGACA	2640
D	2581	TTTACTGCTCAAAACCAAAATCTTGGAAAGAAAGTGGGATGCTTATTTCAAGACA	2640
Q	2641	AGTTACAGCTGTAAACAAAGACATTAATTTGAGATGAGATGCCAAACCTGTATAAA	2700
D	2641	AGTTACAGCTGTAAACAAAGACATTAATTTGAGATGAGATGCCAAACCTGTATAAA	2700
Q	2701	TGTCCTGTATCAATCACTTCTCAAGATTTCTTCAATGAGCTTCACTTTTACAGA	2760
D	2701	TGTCCTGTATCAATCACTTCTCAAGATTTCTTCAATGAGCTTCACTTTTACAGA	2760
Q	2761	ACTCTGTGTGTGAGATGAGACTTGAAGGAGGTGAGGAGAGTGTGAATAGTGCT	2820
D	2761	ACTCTGTGTGTGAGATGAGACTTGAAGGAGGTGAGGAGAGTGTGAATAGTGCT	2820
Q	2821	TCCCTGAGCAGAAATGTCTACATCTTGAACAAACAGATGTAACCTTAAGCTTCC	2880
D	2821	TCCCTGAGCAGAAATGTCTACATCTTGAACAAACAGATGTAACCTTAAGCTTCC	2880
Q	2881	ATTCACTTGTAAATAATTTGATGATGATGATGATGATGATGATGATGATGATGATG	2940
D	2881	ATTCACTTGTAAATAATTTGATGATGATGATGATGATGATGATGATGATGATGATG	2940
Q	2941	TTTAAATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	3000
D	2941	TTTAAATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	3000
Q	3001	TATCTGAAAGAAATATAGAGCAAGACCTGTGCTCAATATATAGGAATGCTTCT	3060
D	3001	TATCTGAAAGAAATATAGAGCAAGACCTGTGCTCAATATATAGGAATGCTTCT	3060
Q	3061	TTAGTCTTCAAGACTATGTTGTGAAAAACAATGAGGGGTCTAATCTCTTAAGAGTAC	3120
D	3061	TTAGTCTTCAAGACTATGTTGTGAAAAACAATGAGGGGTCTAATCTCTTAAGAGTAC	3120
Q	3121	CTTTTATCTTAAAGAAATATGTCCTGAGATTTATAGCACTTTTGAAGAGAGCCAG	3180
D	3121	CTTTTATCTTAAAGAAATATGTCCTGAGATTTATAGCACTTTTGAAGAGAGCCAG	3180
Q	3181	GTAATGAGGTGTGTGCTGAGCAGATGAGAGCAAGAGAGAGATGAGATACATTTG	3240
D	3181	GTAATGAGGTGTGTGCTGAGCAGATGAGAGCAAGAGAGAGATGAGATACATTTG	3240
Q	3241	TGGAGAGAGAAAGATTTCTCAGAGGCTCCCACTGCTAAAGTTTGTGAGATGTT	3300
D	3241	TGGAGAGAGAAAGATTTCTCAGAGGCTCCCACTGCTAAAGTTTGTGAGATGTT	3300

Dh 661 GCGAGGCGATGAGAGATGAGAGAGAGCCCTCTCGGGGGTGTGGCGCTCGGAG 720
Qy 721 CCTCAGTCAAGCGCGCTGCTCAAAAGATCTTCTGTGCAATTGGAACCTCATCGAAC 780
Dh 721 CCTCAGTCAAGCGCGCTGCTCAAAAGATCTTCTGTGCAATTGGAACCTCATCGAAC 780
Qy 781 AGCAGCAGCAGCAGCTGCGAGGCCAAGAAAAGAGATCGAGAGCTGAAGTCAGAGAG 840
Dh 781 AGCAGCAGCAGCAGCTGCGAGGCCAAGAAAAGAGATCGAGAGCTGAAGTCAGAGAG 840
Qy 841 ACACGCTCTCTGCTCGGATGGAAGATGAGAAAGCGGATGCGAGCTGTAAAGAGATA 900
Dh 841 ACACGCTCTCTGCTCGGATGGAAGATGAGAAAGCGGATGCGAGCTGTAAAGAGATA 900
Qy 901 ACAGAGAAAGAAAGCACAAGCTGTTTCAGGGCTATGAAACTGAAAGAGAGAGAAACAG 960
Dh 901 ACAGAGAAAGAAAGCACAAGCTGTTTCAGGGCTATGAAACTGAAAGAGAGAGAAACAG 960
Qy 961 AGCTATCTGAGAAATTTAACTGAGAGTCCAGCGGAGCTTTCGAGACATCCAGATC 1020
Dh 961 AGCTATCTGAGAAATTTAACTGAGAGTCCAGCGGAGCTTTCGAGACATCCAGATC 1020
Qy 1021 TGCCTCCCAAGCCCTTCTCATGTGGCGAGTGGAAAAGGACATAAAGAAATCCCAT 1080
Dh 1021 TGCCTCCCAAGCCCTTCTCATGTGGCGAGTGGAAAAGGACATAAAGAAATCCCAT 1080
Qy 1081 TTGGAAGTACAGAAAGAAAGCTCTGTAAAGCTGGCTCCGATTTTCAAAAGTCA 1140
Dh 1081 TTGGAAGTACAGAAAGAAAGCTCTGTGTAAAGAGCTGGCTCCGAAATTTCAAAAGTCA 1140
Qy 1141 AAACAAAACTCTTAAGCACTCTCTTAATTAAGAGAAACCTGTGTCTTAACTGAAA 1200
Dh 1141 AAACAAAACTCTTAAGCACTCTCTTAATTAAGAGAAACCTGTGTCTTAACTGAAA 1200
Qy 1201 CTGTGTGTAAAGCTGAATTAAGAGAGCAAGAAACCCCAAGAAAGCCCGTCTTCACTG 1260
Dh 1201 CTGTGTGTAAAGCTGAATTAAGAGAGCAAGAAACCCCAAGAAAGCCCGTCTTCACTG 1260
Qy 1261 ACACCCCAAGCAAGCTCTCACTCCCAAGAGGAGCCAGACCCATCCCAAGAGAAAG 1320
Dh 1261 ACACCCCAAGCAAGCTCTCACTCCCAAGAGGAGCCAGACCCATCCCAAGAGAAAG 1320
Qy 1321 CTTTCTCAAGTGAATGAAGATTTGCCGTAACCTTTCACACAGAAATGATTTGTGTC 1380
Dh 1321 CTTTCTCAAGTGAATGAAGATTTGCCGTAACCTTTCACACAGAAATGATTTGTGTC 1380
Qy 1381 GTTGGCACCAAGCTCCCAATCACCGTTACATTAACGGAATCTCTCCAAAGAGAGG 1440
Dh 1381 GTTGGCACCAAGCTCCCAATCACCGTTACATTAACGGAATCTCTCCAAAGAGAGG 1440
Qy 1441 AGACTGTAGAGAGTGTCTGATGCGATCAAGTGTGAGAGAACTTCACTGTGGCTG 1500
Dh 1441 AGACTGTAGAGAGTGTCTGATGCGATCAAGTGTGAGAGAACTTCACTGTGGCTG 1500
Qy 1501 TTCTCTTTGAGAGGACCACTCAGTAGAGCTCTTAAGGAGACCAAAATCTTTCAGACTTT 1560
Dh 1501 TTCTCTTTGAGAGGACCACTCAGTAGAGCTCTTAAGGAGACCAAAATCTTTCAGACTTT 1560
Qy 1561 TGGAGAACTGTGATGACAGTGTGTTTGAAGCGGCAATGCAAACTGAGAGTGAAGA 1620
Dh 1561 TGGAGAACTGTGATGACAGTGTGTTTGAAGCGGCAATGCAAACTGAGAGTGAAGA 1620
Qy 1621 AGAGAAAGAAAGATGAGATTTTCAGAGATCAGGGAACAAAGAAATTTTACAGGACTGC 1680
Dh 1621 AGAGAAAGAAAGATGAGATTTTCAGAGATCAGGGAACAAAGAAATTTTACAGGACTGC 1680
Qy 1681 AGCTCAAGATGTATTAAGAAAGAAAGAAATTCAGGAATCGAGCTGAGATTAACCTCATTTT 1740
Dh 1681 AGCTCAAGATGTATTAAGAAAGAAAGAAATTCAGGAATCGAGCTGAGATTAACCTCATTTT 1740
Qy 1741 TCCCTGAGCCAGATGATGTTGAAAGTTGATGATTAACCCCTTCTTGCCTGTGTAGCAT 1800
Dh 1741 TCCCTGAGCCAGATGATGTTGAAAGTTGATGATTAACCCCTTCTTGCCTGTGTAGCAT 1800

Qy 1801 TTGAGACACCATTAACCAAAATTAATCTCAAGAAATTTGAGCTAACCTGTGTGATGAGC 1860
Dh 1801 TTGAGACACCATTAACCAAAATTAATCTCAAGAAATTTGAGCTAACCTGTGTGATGAGC 1860
Qy 1861 GTAGCCGATTCAGATTTGAGATTCAGAGAAAGCAAACTCACCGGAGCTGTAGAAAT 1920
Dh 1861 GTAGCCGATTCAGATTTGAGATTCAGAGAAAGCAAACTCACCGGAGCTGTAGAAAT 1920
Qy 1921 AGCTGTGTGGCAAGAACCTGTCTTCAATGTTAGATGATGATCCGAGAGAGGC 1980
Dh 1921 AGCTGTGTGGCAAGAACCTGTCTTCAATGTTAGATGATGATCCGAGAGAGGC 1980
Qy 1981 AGAGACCTGTATATGTGACCTTTGTCTCTCAATATGTTATCACTGTGTAATACCTT 2040
Dh 1981 AGAGACCTGTATATGTGACCTTTGTCTCTCAATATGTTATCACTGTGTAATACCTT 2040
Qy 2041 TCATACTTCTTGAATCTTGTGTTCAATTAATCTGTGATTTCAAAAATCTTTTATCGGCT 2100
Dh 2041 TCATACTTCTTGAATCTTGTGTTCAATTAATCTGTGATTTCAAAAATCTTTTATCGGCT 2100
Qy 2101 AATTGTAGTATGAGAGGATGATTTGGGATTTCTTTTCTTTTGGGAAATGGGCTCTC 2160
Dh 2101 AATTGTAGTATGAGAGGATGATTTGGGATTTCTTTTCTTTTGGGAAATGGGCTCTC 2160
Qy 2161 AAGCTAAAGCTATAGATGAGAGATTCAGAAATTCAGGGGCTGTTCTATACATTTGC 2220
Dh 2161 AAGCTAAAGCTATAGATGAGAGATTCAGAAATTCAGGGGCTGTTCTATACATTTGC 2220
Qy 2221 CTATGTTAAAGGGGATTAAGAGGCTCTCTTCAATTAAGATGAGATGAGAGAGCCCT 2280
Dh 2221 CTATGTTAAAGGGGATTAAGAGGCTCTCTTCAATTAAGATGAGATGAGAGAGCCCT 2280
Qy 2281 TCCCTTGAAGCTGTGCTGATGAGCACTCTTTCACCTGTGACACCTCTTATATGTC 2340
Dh 2281 TCCCTTGAAGCTGTGCTGATGAGCACTCTTTCACCTGTGACACCTCTTATATGTC 2340
Qy 2341 GTATAGATTTTAAACCTTAATTAAGAAACAAACCAACCTGACATGAGCTTTAGAGACAG 2400
Dh 2341 GTATAGATTTTAAACCTTAATTAAGAAACAAACCAACCTGACATGAGCTTTAGAGACAG 2400
Qy 2401 AAGAGAAATGACAAGTGAAGCGATGAGAGCAAGCCATCTTCAAGAGTGAAGAAACATCG 2460
Dh 2401 AAGAGAAATGACAAGTGAAGCGATGAGAGCAAGCCATCTTCAAGAGTGAAGAAACATCG 2460
Qy 2461 GAGAGTGTGATGATCTGTCTGAAAAAGATGTTGATTTGAAACTATCTGTGATAC 2520
Dh 2461 GAGAGTGTGATGATCTGTCTGAAAAAGATGTTGATTTGAAACTATCTGTGATAC 2520
Qy 2521 AGTCATGTGGGAAGGAGTGTGGCTGTATTTTTCAGTTAATGAGATTAACATTTTC 2580
Dh 2521 AGTCATGTGGGAAGGAGTGTGGCTGTATTTTTCAGTTAATGAGATTAACATTTTC 2580
Qy 2581 TTTTACTGCTCAAAAACCAAAATCTTTGGAAGAAAGATGGGATGTTAGTTTCAGAAACA 2640
Dh 2581 TTTTACTGCTCAAAAACCAAAATCTTTGGAAGAAAGATGGGATGTTAGTTTCAGAAACA 2640
Qy 2641 AGTTACAGCTGTAAACAAAGCACTTAATTTGGGATGCGATGCCAAAACCTGTATPAA 2700
Dh 2641 AGTTACAGCTGTAAACAAAGCACTTAATTTGGGATGCGATGCCAAAACCTGTATPAA 2700
Qy 2701 TGTCTGTGTATCAATCAATCTTCAAGTATTTCTTCAATGGGCTTCACTCTTTTAGAGA 2760
Dh 2701 TGTCTGTGTATCAATCAATCTTCAAGTATTTCTTCAATGGGCTTCACTCTTTTAGAGA 2760
Qy 2761 ACTCTGTGTGTGGATGAGACTTGAAGAGGTAGGGGAGAGTGTGAATATGATGCT 2820
Dh 2761 ACTCTGTGTGTGGATGAGACTTGAAGAGGTAGGGGAGAGTGTGAATATGATGCT 2820
Qy 2821 TCCCTTGTGTGGCAAAATGTCTAATCTTGAACAAACAGATGTAAATGAGATTTCTCC 2880
Dh 2821 TCCCTTGTGTGGCAAAATGTCTAATCTTGAACAAACAGATGTAAATGAGATTTCTCC 2880

QY	2881	ATTCACTTTGAAAAAATATTTGTAATGTCACATCTTGACCTCTCCCTCCCGTTTG	2940
Db	2881	ATTCACTTTGAAAAAATATTTGTAATGTCACATCTTGACCTCTCCCTCCCGTTTG	2940
QY	2881	ATTCACTTTGAAAAAATATTTGTAATGTCACATCTTGACCTCTCCCTCCCGTTTG	2940
QY	2941	TTAAATATCAGATAGACCTCCAGGCGACCTTGGTCTCAGTGTAAATCCCTATTAC	3000
Db	2941	TTAAATATCAGATAGACCTCCAGGCGACCTTGGTCTCAGTGTAAATCCCTATTAC	3000
QY	3001	TATCTGAAAGAAAAATAGAGCCAGACCTCTGGTCTCAAAATATAGGAATTTGCTTCT	3060
Db	3001	TATCTGAAAGAAAAATAGAGCCAGACCTCTGGTCTCAAAATATAGGAATTTGCTTCT	3060
QY	3061	TTAGCTTCAGAGCATTTGTGTGAACCAAGTAGGGGTCTAATCTCTAGAAAGTAGGGG	3120
Db	3061	TTAGCTTCAGAGCATTTGTGTGAACCAAGTAGGGGTCTAATCTCTAGAAAGTAGGGG	3120
QY	3121	CTTTATCTTAAAGAGATATGTCGCCAGATTATTAGACCTTTTAGAGGAAGCCAAAG	3180
Db	3121	CTTTATCTTAAAGAGATATGTCGCCAGATTATTAGACCTTTTAGAGGAAGCCAAAG	3180
QY	3181	GTAATGAGGGGTGTGGCTGGCCCATCAGTGGAGCACAAGAAAGGAATACATTG	3240
Db	3181	GTAATGAGGGGTGTGGCTGGCCCATCAGTGGAGCACAAGAAAGGAATACATTG	3240
QY	3241	TGGGAAGAGAAAGAAAGTTCCCTCAGGGGCTCCCACTGCTAAAGTTTTTGTGAAGTT	3300
Db	3241	TGGGAAGAGAAAGAAAGTTCCCTCAGGGGCTCCCACTGCTAAAGTTTTTGTGAAGTT	3300
QY	3301	GATCTGTCTCTCCGGAATTTGACCTTTTAAAGGAATATTCTGGCAGCAGATGATATC	3360
Db	3301	GATCTGTCTCTCCGGAATTTGACCTTTTAAAGGAATATTCTGGCAGCAGATGATATC	3360
QY	3361	TTGGATGATCTTGGCTCTTATTTCTCTCTTGTGTGTGTGTGTGTGTGTGTGTGCTA	3420
Db	3361	TTGGATGATCTTGGCTCTTATTTCTCTCTTGTGTGTGTGTGTGTGTGTGTGTGCTA	3420
QY	3421	TGGGTTTCAATTGTAACCTCACCTGCTTAAGAGAGTGGGCTCTCTAATAAGGAACTCGC	3480
Db	3421	TGGGTTTCAATTGTAACCTCACCTGCTTAAGAGAGTGGGCTCTCTAATAAGGAACTCGC	3480
QY	3481	TGTAATACTTCAATTGAGCAGAGATGTAGAAGAAATAGACCTTAATTCACATAGGGCTC	3540
Db	3481	TGTAATACTTCAATTGAGCAGAGATGTAGAAGAAATAGACCTTAATTCACATAGGGCTC	3540
QY	3541	TCATCTCACACCTTAAGAGAGAGATTTCTAGAAAACTGGGCGAGATTTCTTGTCTC	3600
Db	3541	TCATCTCACACCTTAAGAGAGAGATTTCTAGAAAACTGGGCGAGATTTCTTGTCTC	3600
QY	3601	CATCATTTTAAATGTGGCAGCGCTGTCACTTTCTTACTCTTAACCTATGTGATTTCTTC	3660
Db	3601	CATCATTTTAAATGTGGCAGCGCTGTCACTTTCTTACTCTTAACCTATGTGATTTCTTC	3660
QY	3661	GTAACGTGTCCAAAAAGAAAAAGACCAGATCAGTGTCTTTGACTTTGTTCTTTGATCC	3720
Db	3661	GTAACGTGTCCAAAAAGAAAAAGACCAGATCAGTGTCTTTGACTTTGTTCTTTGATCC	3720
QY	3721	CTCAGTTTCTTCTGATTTCAAGCATGTGTGGGTTCTTAATTTTGGTATGATTAGCA	3780
Db	3721	CTCAGTTTCTTCTGATTTCAAGCATGTGTGGGTTCTTAATTTTGGTATGATTAGCA	3780
QY	3781	ATTTAACCAATGTGTGTGTGTGCTTACCCAGAGGGGACCTCCACAGTTTCTGATTAAGT	3840
Db	3781	ATTTAACCAATGTGTGTGTGTGCTTACCCAGAGGGGACCTCCACAGTTTCTGATTAAGT	3840
QY	3841	CTGAGAAGAAATCAGAGGTGTCTATCTGGCCAGATTTAAGTAAGATCTAATTTCTGGTT	3900
Db	3841	CTGAGAAGAAATCAGAGGTGTCTATCTGGCCAGATTTAAGTAAGATCTAATTTCTGGTT	3900
QY	3901	CTCCCTCTCCCGAGACCTCTTATTTTATTTGTCCTCCCTCTCTAGATTAAATTTCTCTTG	3960
Db	3901	CTCCCTCTCCCGAGACCTCTTATTTTATTTGTCCTCCCTCTCTCTAGATTAAATTTCTCTTG	3960
QY	3961	ATTTGACTTTGTTGAAGAGAGTTTGGACAGTATGTAGCAAAAGTTTCCAAGTCAAAAT	4020

Db	3961	ATTTGACTTTGTTGAGAAAGAGGTGGACAGTAGATTAGCAAAAGTTCCAAGTGCAAAATT	4020
Qy	4021	ACAGTGTGTAGAGTGTGGGGGGAAAAATTAGTCTTATTTTTCCTTACATGGGATACACA	4080
Db	4021	ACAGTGTGTAGAGTGTGGGGGGAAAAATTAGTCTTATTTTTCCTTACATGGGATACACA	4080
Qy	4081	CTGTGAATTCATCTTCAACGTGAAGGCCCTGCAGTTCTCTTAAACATAGTGTGTTT	4140
Db	4081	CTGTGAATTCATCTTCAACGTGAAGGCCCTGCAGTTCTCTTAAACATAGTGTGTTT	4140
Qy	4141	TTCTTTAACAAAGTTTAAAGCTAGTGTAAATTAATAAAAAAATGCTTGCTGTCTACT	4200
Db	4141	TTCTTTAACAAAGTTTAAAGCTAGTGTAAATTAATAAAAAAATGCTTGCTGTCTACT	4200
Qy	4201	TCAGCTTGTTTATGCCCCATTTCAATGTGTGTGTGTGTGAATTCATATCACTTTGAT	4260
Db	4201	TCAGCTTGTTTATGCCCCATTTCAATGTGTGTGTGTGTGAATTCATATCACTTTGAT	4260
Qy	4261	ACCATTTCTGATGTGTAAAAATGGTGTCTCTGTAAATATCTTATTAAGAGTTCATTTGA	4320
Db	4261	ACCATTTCTGATGTGTAAAAATGGTGTCTCTGTAAATATCTTATTAAGAGTTCATTTGA	4320
Qy	4321	AATTAACATATGTGCGCTGTAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	4372
Db	4321	AATTAACATATGTGCGCTGTAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	4372

```

RESULT 4
US-09-397-945-79
; Sequence 79, Application US/09397945
; Publication No. US20030065139A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc. et al.
; TITLE OF INVENTION: 95 Human secreted proteins
; FILE REFERENCE: P2027P1
; CURRENT APPLICATION NUMBER: US/09/397,945
; CURRENT FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: PCT/US99/05804
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/078,566
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,576
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,573
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,574
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,579
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,314
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080,312
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/078,578
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,581
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,577
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,563
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,313
; PRIOR FILING DATE: 1998-04-01
; NUMBER OF SEQ ID NOS: 470
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 79
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-397-945-79

Query Match      28.8%; Score 1360; DB 10; Length 1509

```


Best Local Similarity 99.98; Pred. No. 0; Mismatches 0; Indels 2; Gaps 2;
Matches 1500; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 2859 GATGATACCTATGAGTCTTCATCTCATCTTGTAAAAATATTTGTATGTACCATCTT 2918
Db 9 GATGATACCTATGAGTCTTCATCTCATCTTGTAAAAATATTTGTATGTACCATCTT 68
QY 2919 GGTCTCTCCCTCCCGTTTTGTAAAAATATGAGATAGCATCTCCAGGCACTTGTGTC 2978
Db 69 GGTCTCTCCCTCCCGTTTTGTAAAAATATGAGATAGCATCTCCAGGCACTTGTGTC 128
QY 2979 TCACTGTAAATCCCTATTAATCTGAAAGAAAATAGAGCAAGCTTGTGTC 3038
Db 129 TCACTGTAAATCCCTATTAATCTGAAAGAAAATAGAGCAAGCTTGTGTC 188
QY 3039 AATATATAGAAATTCCTCTTCTTGTAGCTTCAGAGCAATGTGTGAAAACAAGAGGGGT 3098
Db 189 AATATATAGAAATTCCTCTTCTTGTAGCTTCAGAGCAATGTGTGAAAACAAGAGGGGT 248
QY 3099 CTATCTCTAGAGAGTAGGGGCTTTTATCTTAAAGAAATATGTCCAGATTTATAG 3158
Db 249 CTATCTCTAGAGAGTAGGGGCTTTTATCTTAAAGAAATATGTCCAGATTTATAG 307
QY 3159 CACTTTAGAGAGAGCAAGGATATGA- GGGTGTGTGCTGGCCATCATGTGAGACAC 3217
Db 308 CACTTTAGAGAGAGCAAGGATATGA- GGGTGTGTGCTGGCCATCATGTGAGACAC 367
QY 3218 GAAGAGAGATGAGATCCATTTGTGGAAGAGAAAGTCTCTCAGGGGCTCCCACT 3277
Db 368 GAAGAGAGATGAGATCCATTTGTGGAAGAGAAAGTCTCTCAGGGGCTCCCACT 427
QY 3278 GCTAAAGTTTTTGTGAGATGTGATCTGTGCTCTGATTTGACCTTTAAAGAAATTA 3337
Db 428 GCTAAAGTTTTTGTGAGATGTGATCTGTGCTCTGATTTGACCTTTAAAGAAATTA 487
QY 3338 TTCTGGAGACATGATGATATCTTGGATGATCTTGTGCTCTTATTTCTCTTTTGT 3397
Db 488 TTCTGGAGACATGATGATATCTTGGATGATCTTGTGCTCTTATTTCTCTTTTGT 547
QY 3398 GT 3457
Db 548 GT 607
QY 3458 GGGCTCTTAAAGAGGAACTGTCTTAACTTCTTGTGAGAGAGATGTGAGAGAAATA 3517
Db 608 GGGCTCTTAAAGAGGAACTGTCTTAACTTCTTGTGAGAGAGATGTGAGAGAAATA 667
QY 3518 GGAATTAACTCACTAGGGGCTCTCATCTCAACCTTAAAGAGAGATTTCTAGAAAAC 3577
Db 668 GGAATTAACTCACTAGGGGCTCTCATCTCAACCTTAAAGAGAGATTTCTAGAAAAC 727
QY 3578 TGGGCCAGATTTCTTGTCTCTCATCATTTTAAATGTGGCAGGCTGTTCAGTTTCTTAC 3637
Db 728 TGGGCCAGATTTCTTGTCTCTCATCATTTTAAATGTGGCAGGCTGTTCAGTTTCTTAC 787
QY 3638 TCTTACCTAGTATATTTCTTGTCTCTCACTCTCAAAAAGAAAAGAAAACCAATCAAGT 3697
Db 788 TCTTACCTAGTATATTTCTTGTCTCTCACTCTCAAAAAGAAAAGAAAACCAATCAAGT 847
QY 3698 CTCCTGACTTGTCTTGTGATCCCTCAGTTTCTTGTGATTTGAGATGTGCGGTTCC 3757
Db 848 CTCCTGACTTGTCTTGTGATCCCTCAGTTTCTTGTGATTTGAGATGTGCGGTTCC 907
QY 3758 TAATTTGGGATATGATTAAGAAATTTAACATTTGTGTGGCCCTTAAAGAGGAGCTC 3817
Db 908 TAATTTGGGATATGATTAAGAAATTTAACATTTGTGTGGCCCTTAAAGAGGAGCTC 967
QY 3818 CCCAGTTTCTGACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 3877
Db 968 CCCAGTTTCTGACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1027
QY 3878 AAGTAGATCTATTTCTTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3937

Db 1028 AAGTAGATCTATTTCTTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1087
QY 3938 TCTTCTAGATTAATTCCTCTTGTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 3997
Db 1088 TCTTCTAGATTAATTCCTCTTGTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1147
QY 3998 AGCAAGTTCCAAAGTGAAGAAATTAAGTGTGTGATGAGTGTGGGGGAAAATTAAGTCTTAT 4057
Db 1148 AGCAAGTTCCAAAGTGAAGAAATTAAGTGTGTGATGAGTGTGGGGGAAAATTAAGTCTTAT 1207
QY 4058 TTTTCCCTAATGAGTATCAACATGATGATTTCAATTTCAATTTCAATTTCAATTTCAATTTCA 4117
Db 1208 TTTTCCCTAATGAGTATCAACATGATGATTTCAATTTCAATTTCAATTTCAATTTCAATTTCA 1267
QY 4118 TCCATAAATGATGATTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4177
Db 1268 TCCATAAATGATGATTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1327
QY 4178 AAAAAATGCTGTCTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4237
Db 1328 AAAAAATGCTGTCTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1387
QY 4238 TGTGTATTCATTAATCTTGTATTAATCTTGTATTAATCTTGTATTAATCTTGTATTAATCTTGT 4297
Db 1388 TGTGTATTCATTAATCTTGTATTAATCTTGTATTAATCTTGTATTAATCTTGTATTAATCTTGT 1447
QY 4298 ATCTTAAAGATTCATTTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4357
Db 1448 ATCTTAAAGATTCATTTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1507
QY 4358 AA 4359
Db 1508 AA 1509
RESULT 5
US-10-653-595-79
; Sequence 79, Application US/10653595
; Publication No. US20040048304A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OR INVENTION: 95 Human secreted proteins
; FILE REFERENCE: P2027P1C1
; CURRENT APPLICATION NUMBER: US/10/653,595
; CURRENT FILING DATE: 2003-09-03
; PRIOR APPLICATION NUMBER: US 09/397945
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: PCT/US99/05804
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/078,566
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,576
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,573
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,574
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,579
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,314
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080,312
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/078,578
; PRIOR FILING DATE: 1998-03-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 470
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 79
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-653-595-79

Query Match 28.8%; Score 1260; DB 17; Length 1509;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1500; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 2859 GATGTACCTAATAGCTTCCATCTTCACTTGTAAAAATTAATTTGATGTGACCATCTT 2918
DB 9 GATGTACCTAATAGCTTCCATCTTCACTTGTAAAAATTAATTTGATGTGACCATCTT 68
QY 2919 GGTCTCTCCCTCCCGTTTGTAAAAATATCAGATAGCATCTCCAGGCCACTTTGTC 2978
DB 69 GGTCTCTCCCTCCCGTTTGTAAAAATATCAGATAGCATCTCCAGGCCACTTTGTC 128
QY 2979 TCAGTGTAGATCCCTATTACTATCTGAAAGAAAAATAGCCAGACTCTGTCTCA 3038
DB 129 TCAGTGTAGATCCCTATTACTATCTGAAAGAAAAATAGCCAGACTCTGTCTCA 188
QY 3039 AATATATAGGAATTTGCTTTCTTTAGCTCAGAGCATATGTGTGAAAAAAGAGGGGT 3098
DB 189 AATATATAGGAATTTGCTTTCTTTAGCTCAGAGCATATGTGTGAAAAAAGAGGGGT 248
QY 3099 CTAACTCTCAGAAAGTATGGGGCTTTTATCTTAAAGAAATATGTCCCGATTATTAG 3158
DB 249 CTAACTCTCAGAAAGTATGGGGCTTTTATCTTAAAGAAATATGTCCCGATTATTAG 307
QY 3159 CACTTTAAGAGGAAGCCAGATATGTA-GGGTGTGTGCTGGCCCATCATGTGAGCAC 3217
DB 308 CACTTTAAGAGGAAGCCAGATATGTAAGGGGTGTGTGCTGGCCCATCATGTGAGCAC 367
QY 3218 GAAGAGGAATGGGATATCCATTTGGGAGAGAAAGAAAGTTCTCAGGGGCTCCCACT 3277
DB 368 GAAGAGGAATGGGATATCCATTTGGGAGAGAAAGAAAGTTCTCAGGGGCTCCCACT 427
QY 3278 GCTAAAGTTTTTGTAGATGTGATCTGTGCTTCTGATTTGACTTTTAAAGAAATTA 3337
DB 428 GCTAAAGTTTTTGTAGATGTGATCTGTGCTTCTGATTTGACTTTTAAAGAAATTA 487
QY 3338 TTCTGGCAGACATGTAATTTCTTGATATCTTGCTCTTATTTCTCTTTGTGT 3397
DB 488 TTCTGGCAGACATGTAATTTCTTGATATCTTGCTCTTATTTCTCTTTGTGT 547
QY 3398 GT 3457
DB 548 GT 607
QY 3458 GGGCTCTCTAATAGGAAGCACTGTGTAACTTCAATTCAGCAAGATGTAGAGAAATA 3517
DB 608 GGGCTCTCTAATAGGAAGCACTGTGTAACTTCAATTCAGCAAGATGTAGAGAAATA 667
QY 3518 GGAATTATTCACATAGGGGCTCTCATCTCAACACTTAAGAGAGATTTCTAGAAAAAC 3577
DB 668 GGAATTATTCACATAGGGGCTCTCATCTCAACACTTAAGAGAGATTTCTAGAAAAAC 727
QY 3578 TGGGGCAGATTTTCTTTGTCTCATCATATTTAAATGTGCAAGGCTTTGAGTTTCTTAC 3637
DB 728 TGGGGCAGATTTTCTTTGTCTCATCATATTTAAATGTGCAAGGCTTTGAGTTTCTTAC 787
QY 3638 TCTTACCTATGATATTTCTTGTGTAAGCTGTCCAAAAAGAAAAAGAACCAATCACTGT 3697
DB 788 TCTTACCTATGATATTTCTTGTGTAAGCTGTCCAAAAAGAAAAAGAACCAATCACTGT 847
QY 3698 CTCTTGAATTGTTCTTTGATATCCCTCAGTTTCTTTGATTTTCAAGATGTGTGGGTTCC 3757
DB 848 CTCTTGAATTGTTCTTTGATATCCCTCAGTTTCTTTGATTTTCAAGATGTGTGGGTTCC 907
QY 3758 TAATTTTGGGTATGATAGCAAAATTTAAACATTTGTGTTGTGCTTACCAAGGGGATC 3817
DB 908 TAATTTTGGGTATGATAGCAAAATTTAAACATTTGTGTTGTGCTTACCAAGGGGATC 967
QY 3818 CCAGATTTCTGACTTGAAGTATGACTGAGAAATCCAGAGGTCTATCTGAGGAGATTT 3877
DB 968 CCAGATTTCTGACTTGAAGTATGACTGAGAAATCCAGAGGTCTATCTGAGGAGATTT 1027

QY 3878 AAGTAAATCTAATTTCCCTGTTCTCCCTCCCTGAGAGACCTAATTTTATTTGCCCC 3937
DB 1028 AAGTAAATCTAATTTCCCTGTTCTCCCTCCCTGAGAGACCTAATTTTATTTGCCCC 1087
QY 3938 TCTTCTAGGTAAATTTCTCTTGTGATTTGACTTTGTGAGAAAGAGGTGTGACAGTAAAT 3997
DB 1088 TCTTCTAGGTAAATTTCTCTTGTGATTTGACTTTGTGAGAAAGAGGTGTGACAGTAAAT 1147
QY 3998 AGCAAGTTTCAAGTCAAAATTTACAGTGTGTTAGAGTGTGGGGGAAAAATTGCTTAT 4057
DB 1148 AGCAAGTTTCAAGTCAAAATTTACAGTGTGTTAGAGTGTGGGGGAAAAATTGCTTAT 1207
QY 4058 TTTTCCCTCAATGAGGTATACACATCTGAAATTTCACTCACTGAAAGCCCTGCAAGTTC 4117
DB 1208 TTTTCCCTCAATGAGGTATACACATCTGAAATTTCACTCACTGAAAGCCCTGCAAGTTC 1267
QY 4118 TCCATAAACAATGTTTGTGTTTCTTTTAAACAAATTTAAGCTAGTAAATTAATTA 4177
DB 1268 TCCATAAACAATGTTTGTGTTTCTTTTAAACAAATTTAAGCTAGTAAATTAATTA 1327
QY 4178 AAAAAATGCTGTGTCTACTTCACTTCACTTGTGTTTATGCTTATGCTTGTCTG 4237
DB 1328 AAAAAATGCTGTGTCTACTTCACTTCACTTGTGTTTATGCTTATGCTTGTCTG 1387
QY 4238 TGTGTAAATCAATCTTTTGTATACATTTCTGATGTAATTTGTTGTCTTGTAAAT 4297
DB 1388 TGTGTAAATCAATCTTTTGTATACATTTCTGATGTAATTTGTTGTCTTGTAAAT 1447
QY 4298 ATCTTAAGAAGTCAATTTGTAATTAATTAATTTGCTGTGTTAAAAAATTTAAAAA 4357
DB 1448 ATCTTAAGAAGTCAATTTGTAATTAATTAATTTGCTGTGTTAAAAAATTTAAAAA 1507
QY 4358 AA 4359
DB 1508 AA 1509

RESULT 6
US-10-641-643-167
; Sequence 167, Application US/10641643
; Publication No. US20040077003A1
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; Susan G. Stuart
; Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
; GENE EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/641,643
; FILING DATE: 14-Aug-2003
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 167:
SEQUENCE CHARACTERISTICS:
LENGTH: 1545 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TBLYN0101
CLONE: 040476
SEQUENCE DESCRIPTION: SEQ ID NO: 167 :
US-10-641-643-167

Query Match 20.7%; Score 903; DB 17; Length 1545;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3426 TTTCATTGTTAACTCCATCTGCTTAGAGAGTGGGCTCTCTATTAAGGAACTGCTGTAA 3485
DB 631 TTTCATTGTTAACTCCATCTGCTTAGAGAGTGGGCTCTCTATTAAGGAACTGCTGTAA 690
QY 3486 ACTTCATGCGAGAGAGATGATAGAGAAATAGACTTAATTCACATAGGGGCTCTCATC 3545
DB 691 ACTTCATGCGAGAGATGATAGAGAAATAGACTTAATTCACATAGGGGCTCTCATC 750
QY 3546 TCACACCTTAAGAGAGAGATTTCTAGAAAATGAGGCAATTTCTTGTCTCATCA 3605
DB 751 TCACACCTTAAGAGAGAGATTTCTAGAAAATGAGGCAATTTCTTGTCTCATCA 810
QY 3606 TTTTAAATGTCGAGGCTGTCAGTTTCTTACTCTTACTATGATGATATTTCTTGTAAC 3665
DB 811 TTTTAAATGTCGAGGCTGTCAGTTTCTTACTCTTACTATGATGATATTTCTTGTAAC 870
QY 3666 GTGTCCAAAAAGAAAAAGCAATCAATGCTCTGACTTTGTTGATGCCCTCAG 3725
DB 871 GTGTCCAAAAAGAAAAAGCAATCAATGCTCTGACTTTGTTGATGCCCTCAG 930
QY 3726 TTTCTTCTGATTTACAGATGTCGAGGCTCTTAATTTGGGATGATAGCAAAATTTA 3785
DB 931 TTTCTTCTGATTTACAGATGTCGAGGCTCTTAATTTGGGATGATAGCAAAATTTA 990
QY 3786 ACCATTGTGTTGTGCTTACCCCAAGGAGCTCCCACTTTCTGAATTGAAGTACAG 3845
DB 991 ACCATTGTGTTGTGCTTACCCCAAGGAGCTCCCACTTTCTGAATTGAAGTACAG 1050
QY 3846 AAGAATCCAGAGTGTATCTGCGCCAGATTTAAGTATTTCTTCTGTTCTCCC 3905
DB 1051 AAGAATCCAGAGTGTATCTGCGCCAGATTTAAGTATTTCTTCTGTTCTCCC 1110
QY 3906 TCTCCCTGAGAGCTCTTATTTTATGTCCTCTCTAGGTTAATCTGCTTGAATTTG 3965
DB 1111 TCTCCCTGAGAGCTCTTATTTTATGTCCTCTCTAGGTTAATCTGCTTGAATTTG 1170
QY 3966 ACTTGTGTAAGAGAGATGTCAGTATTAAGCAAGTTCACAAAGTCAAAATTAAGT 4025
DB 1171 ACTTGTGTAAGAGAGATGTCAGTATTAAGCAAGTTCACAAAGTCAAAATTAAGT 1230
QY 4026 GTGTAGAGTGTGGGGGAAAAATTAAGTCTTATTTTCCCTACATGGAATCAACACTGTG 4085
DB 1231 GTGTAGAGTGTGGGGGAAAAATTAAGTCTTATTTTCCCTACATGGAATCAACACTGTG 1290
QY 4086 AATTCATCTTCAACTGAGAGGCTGAGTTCCTTAACCATAGTGTGTTGTTTCTT 4145
DB 1291 AATTCATCTTCAACTGAGAGGCTGAGTTCCTTAACCATAGTGTGTTGTTTCTT 1350
QY 4146 TAACAAGTTTAAAGTATGATTAATAATTAATAATTAATAATTAATAATTAATAATTA 4205
DB 1351 TAACAAGTTTAAAGTATGATTAATAATTAATAATTAATAATTAATAATTAATAATTA 4265
QY 4206 TTTGTTTAAAGCCATTTCAATTTGTTGTCGTGTGATTAATCAATTTTGAATACAT 4265
DB 1411 TTTGTTTAAAGCCATTTCAATTTGTTGTCGTGTGATTAATCAATTTTGAATACAT 1470

QY 4266 TTCTGATGCTAAATTTGTTGTTCTTGTAAATCTTATTAAGAGTTCATTTGAATTA 4325
DB 1471 TTCTGATGCTAAATTTGTTGTTCTTGTAAATCTTATTAAGAGTTCATTTGAATTA 1530
QY 4326 ACT 4328
DB 1531 ACT 1533

RESULT 7

US-09-925-302-212
Sequence 212, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OR INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 212
LENGTH: 1529
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-302-212

Query Match 19.1%; Score 837; DB 9; Length 1529;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 837; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 617 GCCACCCCTGCGCCCAACCGCCAGACCCCTGCGGCGCCAGCGAGGCGAGTGA 676
DB 2 GCCACCCCTGCGCCCAACCGCCAGACCCCTGCGGCGCCAGCGAGGCGAGTGA 61
QY 677 GAGTATGAGAAAGAGCCCTCTCGGGGAGTGTGCGGCTCGGAGCTTCAGTACGCGCC 736
DB 62 GAGTATGAGAAAGAGCCCTCTCGGGGAGTGTGCGGCTCGGAGCTTCAGTACGCGCC 121
QY 737 CTGCTCAACAGATCTTCTGCTGCAATTGACCTCATGCAACAGAGAGAGAGT 796
DB 122 CTGCTCAACAGATCTTCTGCTGCAATTGACCTCATGCAACAGAGAGAGT 181
QY 797 GCAAGCCCAAGAAAGAGATCGAGAGTGAAGTCAGAGAGACAGCTCTTGTCTG 856
DB 182 GCAAGCCCAAGAAAGAGATCGAGAGTGAAGTCAGAGAGACAGCTCTTGTCTG 241
QY 857 GATTGAACGTATGAAAGGCGGATGCACTGTGTAAAGAGATTAAGAAAGAAAGCA 916
DB 242 GATTGAACGTATGAAAGGCGGATGCACTGTGTAAAGAGATTAAGAAAGAAAGCA 301
QY 917 CAAGCTGTTCAAGGCTATGAATCTGAAGAGAGAGAAACAGAGTATCTGAATAAT 976
DB 302 CAAGCTGTTCAAGGCTATGAATCTGAAGAGAGAGAAACAGAGTATCTGAATAAT 361
QY 977 TAAATGAGATGCGAGCGAGCTTTCCAGACATCCAGACTCTGCTCCAGCCCTT 1036
DB 362 TAAATGAGATGCGAGCGAGCTTTCCAGACATCCAGACTCTGCTCCAGCCCTT 421
QY 1037 CTGATGTGGCGGAGTGAAGAGGACATAAAGAAATCCCATTTGGAAGTACAGAA 1096
DB 422 CTGATGTGGCGGAGTGAAGAGGACATAAAGAAATCCCATTTGGAAGTACAGAA 481
QY 1097 AAGAGCTCTGTTAAAGAGTGTGCTCTGAATTTTCAAAAGTCAAAACAAACTCTTA 1156
DB 482 AAGAGCTCTGTTAAAGAGTGTGCTCTGAATTTTCAAAAGTCAAAACAAACTCTTA 541
QY 1157 GCACTCTCTATTAAAGAGAACCTGTGTTCTTATCTGAACCTGTTGTAAAGCTGA 1216

Db 542 GCACTCTCTATTAAAGAGAAACCTGTGTTCTTATCTGAAACGTGTTTAAGTGA 601
Qy 1217 ATTGAGAGCCCAAGAAACCCAGAAAAGCCCGGTCTTCACTGAGACACCCCAAGACT 1276
Db 602 ATTGAGAGCCCAAGAAACCCAGAAAAGCCCGGTCTTCACTGAGACACCCCAAGACT 661
Qy 1277 CTCCTCTCTCTAAAGAGAGCCAGCAACCAATCCAGAGAGAAAGCCTTCTCAAGTAGAT 1336
Db 662 CTCCTCTCTCTAAAGAGAGCCAGCAACCAATCCAGAGAGAAAGCCTTCTCAAGTAGAT 721
Qy 1337 AGAAGATTTGCGGTACCTTTCCACACAGAAATGTATTTGTGCTGTGGACAGCCTCC 1396
Db 722 AGAAGATTTGCGGTACCTTTCCACACAGAAATGTATTTGTGCTGTGGACAGCCTCC 781
Qy 1397 CCATACCGGTACCATTAACGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG 1453
Db 782 CCATACCGGTACCATTAACGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG 838

RESULT 8

US-09-925-302-212
; Sequence 212, Application US/09925302
; Publication No. US20030064072A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 212
; LENGTH: 1529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-302-212

Query Match 19.1%; Score 837; DB 10; Length 1529;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 837; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 617 GCCACCCCTCGCGCCACCGCACCGCCGGAGACCTTGGCGCGCAGCGAGGGCAGATTGAA 676
Db 2 GCCACCCCTCGCGCCACCGCACCGCCGGAGACCTTGGCGCGCAGCGAGGGCAGATTGAA 61
Qy 677 GAGTATGAGAGAGAGCCCTCTCGGGGGTGTGGCGGCTCGGAGCCTCAAGTACGCCGC 736
Db 62 GAGTATGAGAGAGAGCCCTCTCGGGGGTGTGGCGGCTCGGAGCCTCAAGTACGCCGC 121
Qy 737 CTGCGCTCAACACATCTCTGTGCAATGTGACCTTCATGAAACAGAGAGAGAGAGAGCT 796
Db 122 CTGCGCTCAACACATCTCTGTGCAATGTGACCTTCATGAAACAGAGAGAGAGAGCT 181
Qy 797 GCAGGCCAAGAAAGAGATCGAGAGCTGAAGTCAAGAGAGACACGCTCTTGTGCTCG 856
Db 182 GCAGGCCAAGAAAGAGATCGAGAGCTGAAGTCAAGAGAGACACGCTCTTGTGCTCG 241
Qy 857 GATTGAACGTATGAAAAGCGGATGCACTGTGTAAGAGAGATTAAGAGAAAGAGCA 916
Db 242 GATTGAACGTATGAAAAGCGGATGCACTGTGTAAGAGATTAAGAGAAAGAGCA 301
Qy 917 CAAGCGTTTCAAGGCTATGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 976
Db 302 CAAGCGTTTCAAGGCTATGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
Qy 977 TAAAGTGAAGTGCAGGCGGAGCTTTCAGAGACATCCAGACTCTGCTCCAGAGCCTT 1036
Db 362 TAAAGTGAAGTGCAGGCGGAGCTTTCAGAGACATCCAGACTCTGCTCCAGAGCCTT 421

Qy 1037 CTCATGTGGGCGGAGTGGAGAAAGGACATTAAGAGAAATCCCATTTGSAAGTACAGAAAG 1096
Db 422 CTCATGTGGGCGGAGTGGAGAAAGGACATTAAGAGAAATCCCATTTGSAAGTACAGAAAG 481
Qy 1097 AAAAGTCTCTGTTAAAGAAAGCTGCTCTGTAATTTTCAAAAGTCAAAACAAAATCTCTAA 1156
Db 482 AAAAGTCTCTGTTAAAGAAAGCTGCTCTGTAATTTTCAAAAGTCAAAACAAAATCTCTAA 541
Qy 1157 GCACTCTCTATTAAAGAGAAACCTGTGCTTCTTATCTGAAACTGTTTGTAAAGTGA 1216
Db 542 GCACTCTCTATTAAAGAGAAACCTGTGCTTCTTATCTGAAACTGTTTGTAAAGTGA 601
Qy 1217 ATTGAGAGCCCAAGAAACCCAGAAAAGCCCGGTCTTCACTGAGACACCCCAAGACT 1276
Db 602 ATTGAGAGCCCAAGAAACCCAGAAAAGCCCGGTCTTCACTGAGACACCCCAAGACT 661
Qy 1277 CTCCTCTCTCTAAAGAGAGCCAGCACCCATCCCAAGAGAAAGCCTTCTCAAGTAGAT 1336
Db 662 CTCCTCTCTCTAAAGAGAGCCAGCACCCATCCCAAGAGAAAGCCTTCTCAAGTAGAT 721
Qy 1337 AGAAGATTTGCGGTACCTTTCCACACAGAAATGTATTTGTGCTGTGGACAGCCTCC 1396
Db 722 AGAAGATTTGCGGTACCTTTCCACACAGAAATGTATTTGTGCTGTGGACAGCCTCC 781
Qy 1397 CCATACCGGTACCATTAACGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG 1453
Db 782 CCATACCGGTACCATTAACGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG 838

RESULT 9

US-10-641-643-322

; Sequence 322, Application US/10641643

; Publication No. US20040077003A1

; GENERAL INFORMATION:

; APPLICANT: Cocks, Benjamin G.

; Jeffrey J. Seilhamer

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL

; GENE EXPRESSION

; NUMBER OF SEQUENCES: 1508

; CORRESPONDENCE ADDRESS:

; ADDRESS: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/641,643

; FILING DATE: 14-Aug-2003

; CLASSIFICATION: <Unknown>

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Zeller, Karen J.

; REGISTRATION NUMBER: 37,071

; REFERENCE/DOCKET NUMBER: PA-0001 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 855-0555

; TELEFAX: (650) 845-4166

; INFORMATION FOR SEQ ID NO: 322:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1290 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

LIBRARY: TLYMNOR01
CLONE: 140704
SEQUENCE DESCRIPTION: SEQ ID NO: 322
US-10-641-643-322

Query Match 11.1%; Score 487; DB 17; Length 1290;

Best Local Similarity 99.8%; Pred. No. 1,66-228;
Matches 607; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
QY 847 TCCCTGCTCGATGAACTATGATGAAAGCGGATGACGTGTAAAGATTAACAGAGA 906
DB 1 TCCTTGTCTCGATGAACTATGATGAAAGCGGATGACGTGTAAAGATTAACAGAGA 60
QY 907 AAGAAAGGCAAGAGCTGTTTCAAGGCTATGAACTGAAGAGAGAGAGAAACAGAGCTAT 966
DB 61 AAGAAAGGCAAGAGCTGTTTCAAGGCTATGAACTGAAGAGAGAGAGAAACAGAGCTAT 120
QY 967 CTGAGAAATTTAACTGAGATGCCAGCGGAGCTTTCCGAGACATCCCAAGCTGCTTC 1026
DB 121 CTGAGAAATTTAACTGAGATGCCAGCGGAGCTTTCCGAGACATCCCAAGCTGCTTC 180
QY 1027 CCAAGCCCTTCTCATGTGGGGGAGT-GGAAAGGAGCATAAAGGAAATCCCATTTTGA 1085
DB 181 CCAAGCCCTTCTCATGTGGGGGAGTGGGAGAGGAGCATAAAGGAAATCCCATTTTGA 240
QY 1086 AGTACAGAAAGAAAGACTCTGTTTAAAGCTGCTCTGAAATTTTCAAAAGTCAAAACA 1145
DB 241 AGTACAGAAAGAAAGACTCTGTTTAAAGCTGCTCTGAAATTTTCAAAAGTCAAAACA 300
QY 1146 AAAAAGCTTCTAGACCTCTCTATTAAAGAGAAACCTGTGTCTTATCTGAAATGTT 1205
DB 301 AAAAAGCTTCTAGACCTCTCTATTAAAGAGAAACCTGTGTCTTATCTGAAATGTT 360
QY 1206 TGTAAAGCTGTAATTGAGAGAGCAAGAAACCCAGAAAGCCCGGTCTTTCAGTGAAC 1265
DB 361 TGTAAAGCTGTAATTGAGAGAGCAAGAAACCCAGAAAGCCCGGTCTTTCAGTGAAC 420
QY 1266 CCACCAAGACTCTCCACCTCCCAAAAGGAGCCAGACCCATCCCAAGAGAGAAAGCTTC 1325
DB 421 CCACCAAGACTCTCCACCTCCCAAAAGGAGCCAGACCCATCCCAAGAGAGAAAGCTTC 480
QY 1326 TCAAGTAGATAGAAAGATTTGCGGTACCTTTCCACAGAAATGATTTGTGTGTTG 1385
DB 481 TCAAGTAGATAGAAAGATTTGCGGTACCTTTCCACAGAAATGATTTGTGTGTTG 540
QY 1386 CACCAAGCTCCCATCATCCGTTACCTTACCTTACGGAATCTCTCCAAAGAGAGAGACT 1445
DB 541 CACCAAGCTCCCATCATCCGTTACCTTACCTTACGGAATCTCTCCAAAGAGAGAGACT 600
QY 1446 GTAGCAAG 1453
DB 601 GTAGCAAG 608
```

RESULT 10
US-10-242-535A-51339
Sequence 51339, Application US/10242535A
Publication No. US20040013663A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28

NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 51339
LENGTH: 600
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (31)..(31)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (468)..(468)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: (489)
LOCATION: (489)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (520)..(520)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (589)..(589)
OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-51339

Query Match 10.0%; Score 436; DB 17; Length 600;
Best Local Similarity 100.0%; Pred. No. 2e-203;
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 3548 ACACTTAAAGAGAGATTTCTAGAAAGCTGGGCGAGATTTCTTGTCTCATCAT 3607
DB 32 ACACTTAAAGAGAGATTTCTAGAAAGCTGGGCGAGATTTCTTGTCTCATCAT 91
QY 3608 TTAATGTGAGAGCTGTTCTAGATTTCTTACTTCTATGATATTTCTGTAACT 3667
DB 92 TTAATGTGAGAGCTGTTCTAGATTTCTTACTTCTATGATATTTCTGTAACT 151
QY 3668 GTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3727
DB 152 GTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 211
QY 3728 TCTTCTGATTTAGAGATGAGTGGGTTCTTAATTTGGGTAAGTAACTAAATTAAC 3787
DB 212 TCTTCTGATTTAGAGATGAGTGGGTTCTTAATTTGGGTAAGTAACTAAATTAAC 271
QY 3788 CATTTGTTTGTGCTCCCTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3847
DB 272 CATTTGTTTGTGCTCCCTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 331
QY 3848 GAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3907
DB 332 GAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 391
QY 3908 TCCCTAGAGAGCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3967
DB 392 TCCCTAGAGAGCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 451
QY 3968 TTTGTTGAGAGAGAG 3983
DB 452 TTTGTTGAGAGAGAG 467
```

RESULT 11
US-10-085-783A-51339
Sequence 51339, Application US/10085783A
Publication No. US20040037841A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis

```
FILE REFERENCE: 4231/2002
CURRENT APPLICATION NUMBER: US/10/085,783A
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 51339
LENGTH: 600
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (31)..(31)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (468)..(468)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (489)..(489)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (520)..(520)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (589)..(589)
OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-51339

Query Match
Best Local Similarity 10.0%; Score 436; DB 17; Length 600;
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3548 ACACCTTAAGAGAGATTTCTAGAAAACTGGGCCAGATTTTCTTGTCTGCATCAT 3607
32 ACACCTTAAGAGAGATTTCTAGAAAACTGGGCCAGATTTTCTTGTCTGCATCAT 91
3608 TTAATGTGAGAGCTGTTCAGTTTCTTACTTACCTATGTATTTCTTGTAACGT 3667
92 TTAATGTGAGAGCTGTTCAGTTTCTTACTTACCTATGTATTTCTTGTAACGT 151
3668 GTCCAAAAAGAAAAAGAACCCATCACTGTCTCTTGACTTTTCTTGTATCCCTCAG 3727
152 GTCCAAAAAGAAAAAGAACCCATCACTGTCTCTTGACTTTTCTTGTATCCCTCAG 211
3728 TCTTCTTGATTTGAGCATGTGTGCGGGTTCTTAATTTTGGGATGAGTTAGCAATT 3787
212 TCTTCTTGATTTGAGCATGTGTGCGGGTTCTTAATTTTGGGATGAGTTAGCAATT 211
3788 CATTTGTGTGTGCTCCCTACCCAGAGGACTCCAGTTTCTGACTTGAAGTAGAGAA 3847
272 CATTTGTGTGTGCTCCCTACCCAGAGGACTCCAGTTTCTGACTTGAAGTAGAGAA 331
3848 GAATTCACGAGGTCTATCTGGCCAGATTAAAGATTTCTATTTCTTGAGTTCTCC 3907
332 GAATTCACGAGGTCTATCTGGCCAGATTAAAGATTTCTATTTCTTGAGTTCTCC 391
3908 TCCCTGAGAGCTCTATTTTATTTGCTCCCTCTTGAAGTTATTCCTCTTGAATT 3967
392 TCCCTGAGAGCTCTATTTTATTTGCTCCCTCTTGAAGTTATTCCTCTTGAATT 451
3968 TTTGTTGAGAGAGG 3983
452 TTTGTTGAGAGAGG 467
```

```
RESULT 12
US-10-242-535A-58053
Sequence 58053; Application US/10242535A
Publication No. US20040013663A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Iwew, C. C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 58053
LENGTH: 451
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3)..(3)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (6)..(6)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (417)..(417)
OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-58053

Query Match
Best Local Similarity 9.4%; Score 410; DB 17; Length 451;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3875 TTTAAGTATTTCTATTTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3934
7 TTTAAGTATTTCTATTTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 66
3935 CCTCTCTAGGTTAATTCCTTGAATTTGAATTTGAAGAGAGGTTGACAGTAG 3994
67 CCTCTCTAGGTTAATTCCTTGAATTTGAATTTGAAGAGAGGTTGACAGTAG 126
3995 ATTAGCAAGTTCCAGTGCATAAATTACAGTGTGTAGAGTGTGGGGGAAATTAG 4054
127 ATTAGCAAGTTCCAGTGCATAAATTACAGTGTGTGTAGAGTGTGGGGGAAATTAG 186
4055 TATTTTCCCTACATGGATCAACACTGTGAATTCAATCTTCACTGAAGCCCTG 4114
187 TATTTTCCCTACATGGATCAACACTGTGAATTCAATCTTCACTGAAGCCCTG 246
4115 TTCTCTTAAACATAGTGTGTCTTTCTTTTAAACAAAGTTTAAAGTAGTTAAT 4174
247 TTCTCTTAAACATAGTGTGTCTTTCTTTTAAACAAAGTTTAAAGTAGTTAAT 306
4175 TAAATAAATTTGCTGTCTCTACTTCACTTGTATTTATGCGCAATTTCAATTTG 4234
307 TAAATAAATTTGCTGTCTCTACTTCACTTGTATTTATGCGCAATTTCAATTTG 366
4235 CTGTGTGTATTTCAATTTTGAATCAATTTCTGATGTGTAATTTGG 4284
367 CTGTGTGTATTTCAATTTTGAATCAATTTCTGATGTGTAATTTGG 416
```

```

RESULT 13
US-10-085-783A-58053
; Sequence 58053, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: Liaw, C.C.
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.1.2
; SEQ ID NO 58053
; LENGTH: 451
; TYPE: DNA
; ORGANISM: Human
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (3)..(3)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6)..(6)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (417)..(417)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-58053

Query Match          9.4%; Score 410; DB 17; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.3e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3875 TTTAAGTAGATTCAATTTCCCTGGTCCCTCCCTCCGAGAGACCTCTATTTATTTAGTC 3994
DB 7 TTTAAGTAGATTCAATTTCCCTGGTCCCTCCCTCCGAGAGACCTCTATTTATTTAGTC 66
QY 3935 CCCCTCTTAGGTTAAATCTCCTTGGATTGACCTTGTGTGAGAGAGGTTGACAGTAG 3994
DB 67 CCCCTCTTAGGTTAAATCTCCTTGGATTGACCTTGTGTGAGAGAGGTTGACAGTAG 126
QY 3995 ATTAGCAAAAGTCCAAAGTCAAAATTAACAGTGTGTTAGAGTGTGGGGGAAAATTAGTCT 4054
DB 127 ATTAGCAAAAGTCCAAAGTCAAAATTAACAGTGTGTTAGAGTGTGGGGGAAAATTAGTCT 186
QY 4055 TATTTTTCCTACATGAGGATACACACGTGTGAATCAATCTTCAACTGAAGCCCTGCAG 4114
DB 187 TATTTTTCCTACATGAGGATACACACGTGTGAATCAATCTTCAACTGAAGCCCTGCAG 246
QY 4115 TTTCCTCTAAACATGAGTGTGTTTTCTTAAACAAAGTTAAGCTAGTGTATATAAT 4174
DB 247 TTTCCTCTAAACATGAGTGTGTTTTCTTAAACAAAGTTAAGCTAGTGTATATAAT 306
QY 4175 TAAAAAAAATGCTGTCTGTCTACTGACTGCTTGTGTTATGCCCATTCAATATTTGTTGT 4234
DB 307 TAAAAAAAATGCTGTCTGTCTACTGACTGCTTGTGTTATGCCCATTCAATATTTGTTGT 366
QY 4235 CTGTGTTGATATCTAATCTTTGATATACATTTCTGATGCTGTAATAATTGG 4284
DB 367 CTGTGTTGATATCTAATCTTTGATATACATTTCTGATGCTGTAATAATTGG 416

RESULT 14
US-09-918-995-30064
; Sequence 30064, Application US/09918995

```

```

; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30064
; LENGTH: 365
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-30064

Query Match      8.2%; Score 357; DB 10; Length 365;
Best Local Similarity 100.0%; Pred. No. 1.6e-164;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      919   AGCTGTTTCAGGGCGTATGATAAACTGAAGAGAGAAGAAAACAGACTATCTGAAAAATTTA    978
Db       9   AGCTGTTCAGGGCGTATGATAAATCTGAAGAGAGAAGAAAACAGACTATCTGAAAAATTTA    68

Qy      979   AACTGAGAGTGCCAGCCCGGAGGCTTTTCCGAGACATCTCCAGACTCTGCCTCCCAAGCCTTCT    1038
Db       69   AACTGAGAGTGCCAGCCCGGAGGCTTTTCCGAGACATCTCCAGACTCTGCCTCCCAAGCCTTCT    128

Qy      1039  CATGTGGCGCGAGTGGAAGAAAGGACAATAAAGAAATCCCATTGTTGAAAGTAGACAGAAAGAA    1098
Db      129   CATGTGGCGCGAGTGGAAGAAAGGACAATAAAGAAATCCCATTGTTGAAAGTAGACAGAAAGAA    188

Qy      1099  AGACTCCCGTTAAAAAAGCTGSGCTCCTGAATTTTCAAAAAGTCAAAAACAAAACCTCTAAGC    1158
Db      189   AGACTCCCGTTAAAAAAGCTGSGCTCCTGAATTTTCAAAAAGTCAAAAACAAAACCTCTAAGC    248

Qy      1159  ACTCTCTATTAAAGAGAACCCCTGTGGTTCCTTATCTGAAACTGTTTGTAAACGTGAT      1218
Db      249   ACTCTCTATTAAAGAGAACCCCTGTGGTTCCTTATCTGAAACTGTTTGTAAACGTGAT      308

Qy      1219  TGAGGAGGCCAAGAAACCACAGAAAAAGCCCCGCTCTTCAGTGAACACCCACCAAGAC    1275
Db      309   TGAGGAGGCCAAGAAACCACAGAAAAAGCCCCGCTCTTCAGTGAACACCCACCAAGAC    365

RESULT 15
US-10-242-535A-8833
; Sequence 8833, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8833
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-8833
```

Query Match 6.4%; Score 281; DB 17; Length 400;
Best Local Similarity 100.0%; Pred. No. 4e-127;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3949 AATTCCTTGAATTTGAGTGTGAGAGAGGTTGACAGTAGATTAAGCAAGTCC 4008
DB 69 AATTCCTTGAATTTGAGTGTGAGAGAGGTTGACAGTAGATTAAGCAAGTCC 128
QY 4009 AAGTGAATAATTAACAGTGTGTTAGAGTGTGGGGGAAAATTAAGTCTTATTTTCCCTACA 4068
DB 129 AAGTGAATAATTAACAGTGTGTTAGAGTGTGGGGGAAAATTAAGTCTTATTTTCCCTACA 188
QY 4069 TGGGATACAACTGTCGATTTCAATCTTCACTGGAAGCCCTGACGTTCTCTAAACAT 4128
DB 189 TGGGATACAACTGTCGATTTCAATCTTCACTGGAAGCCCTGACGTTCTCTAAACAT 248
QY 4129 AGTGTGTTGTTTCTTTTAACAAAAGTTTAAGTGTATTAATTAATAAAAAATTTGCT 4188
DB 249 AGTGTGTTGTTTCTTTTAACAAAAGTTTAAGTGTATTAATTAATAAAAAATTTGCT 308
QY 4189 TGTCTGTACTTACAGCTTTGTTTAAGCCCATTTCAATT 4229
DB 309 TGTCTGTACTTACAGCTTTGTTTAAGCCCATTTCAATT 349

RESULT 16
US-10-085-783A-8833
; Sequence 8833, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Ilew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8833
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-8833

Query Match 6.4%; Score 281; DB 17; Length 400;
Best Local Similarity 100.0%; Pred. No. 4e-127;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3949 AATTCCTTGAATTTGAGTGTGAGAGAGGTTGACAGTAGATTAAGCAAGTCC 4008
DB 69 AATTCCTTGAATTTGAGTGTGAGAGAGGTTGACAGTAGATTAAGCAAGTCC 128
QY 4009 AAGTGAATAATTAACAGTGTGTTAGAGTGTGGGGGAAAATTAAGTCTTATTTTCCCTACA 4068
DB 129 AAGTGAATAATTAACAGTGTGTTAGAGTGTGGGGGAAAATTAAGTCTTATTTTCCCTACA 188
QY 4069 TGGGATACAACTGTCGATTTCAATCTTCACTGGAAGCCCTGACGTTCTCTAAACAT 4128
DB 189 TGGGATACAACTGTCGATTTCAATCTTCACTGGAAGCCCTGACGTTCTCTAAACAT 248
QY 4129 AGTGTGTTGTTTCTTTTAACAAAAGTTTAAGTGTATTAATTAATAAAAAATTTGCT 4188
DB 249 AGTGTGTTGTTTCTTTTAACAAAAGTTTAAGTGTATTAATTAATAAAAAATTTGCT 308
QY 4189 TGTCTGTACTTACAGCTTTGTTTAAGCCCATTTCAATT 4229

DB 309 TGTCTGTACTTACAGCTTTGTTTAAGCCCATTTCAATT 349

RESULT 17
US-10-425-115-150265
; Sequence 150265, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 150265
; LENGTH: 273
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_68569C.1
US-10-425-115-150265

Query Match 5.9%; Score 260; DB 18; Length 273;
Best Local Similarity 100.0%; Pred. No. 8.6e-117;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3791 TGTGTTGTCCTTACCCAGGGGACATCCAGTTCTGAAGTAGACAGGAAGA 3850
DB 14 TGTGTTGTCCTTACCCAGGGGACATCCAGTTCTGAAGTAGACAGGAAGA 73
QY 3851 TCCACGAGGTGATCTGAGCCAGATTTAGATTTCTTCTGTTCTCTCTCC 3910
DB 74 TCCACGAGGTGATCTGAGCCAGATTTAGATTTCTTCTGTTCTCTCTCC 133
QY 3911 CTGAGGACCTCTTATTTTATGTCCTCTCTTCTAGGTTAATTCCTTGAATTTGACCTT 3970
DB 134 CTGAGGACCTCTTATTTTATGTCCTCTCTTCTAGGTTAATTCCTTGAATTTGACCTT 193
QY 3971 GTTGAAGAAGGTTGACAGTAGATTAAGTCCAAAGTCCAAATTAAGAGTGCTT 4030
DB 194 GTTGAAGAAGGTTGACAGTAGATTAAGTCCAAAGTCCAAATTAAGAGTGCTT 253
QY 4031 AGAGTGTGGGGGAAAATTA 4050
DB 254 AGAGTGTGGGGGAAAATTA 273

RESULT 18
US-10-242-535A-43646
; Sequence 43646, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Ilew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 43646

LENGTH: 245
TYPE: DNA
ORGANISM: Human
US-10-242-535A-43646

Query Match 5.5%; Score 239; DB 17; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.9e-106;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2471 AGATACTGCTGTAAGAGATGTTGTCATTTGAACTATTCGTGTGATACAGCATGTGG 2530
DB 1 AGATACTGCTGTAAGAGATGTTGTCATTTGAACTATTCGTGTGATACAGCATGTGG 60
QY 2531 GAAGGATGTTGGCTGTGATTAATTTTTCAGTTAATGATTAACAATTTCTTTACTGCTC 2590
DB 61 GAAGGATGTTGGCTGTGATTAATTTTTCAGTTAATGATTAACAATTTCTTTACTGCTC 120
QY 2591 AAAAACCAAAATCTTTGAAAAAGAAAGTGGGATGTTAGTTTCAGAACAGTTACAGCT 2650
DB 121 AAAAACCAAAATCTTTGAAAAAGAAAGTGGGATGTTAGTTTCAGAACAGTTACAGCT 180
QY 2651 GTAAACAAAAGCACTTAGTATTTTGGATGGCATGCCAAAACCTGTATTAATGTCCTTGT 2709
DB 181 GTAAACAAAAGCACTTAGTATTTTGGATGGCATGCCAAAACCTGTATTAATGTCCTTGT 239

RESULT 19
US-10-085-783A-43646
Sequence 43646; Application US/10085783A
Publication No. US20040037841A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Ilew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2002
CURRENT APPLICATION NUMBER: US/10/085,783A
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 43646
LENGTH: 245
TYPE: DNA
ORGANISM: Human
US-10-085-783A-43646

Query Match 5.5%; Score 239; DB 17; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.9e-106;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2471 AGATACTGCTGTAAGAGATGTTGTCATTTGAACTATTCGTGTGATACAGCATGTGG 2530
DB 1 AGATACTGCTGTAAGAGATGTTGTCATTTGAACTATTCGTGTGATACAGCATGTGG 60
QY 2531 GAAGGATGTTGGCTGTGATTAATTTTTCAGTTAATGATTAACAATTTCTTTACTGCTC 2590
DB 61 GAAGGATGTTGGCTGTGATTAATTTTTCAGTTAATGATTAACAATTTCTTTACTGCTC 120
QY 2591 AAAAACCAAAATCTTTGAAAAAGAAAGTGGGATGTTAGTTTCAGAACAGTTACAGCT 2650
DB 121 AAAAACCAAAATCTTTGAAAAAGAAAGTGGGATGTTAGTTTCAGAACAGTTACAGCT 180
QY 2651 GTAAACAAAAGCACTTAGTATTTTGGATGGCATGCCAAAACCTGTATTAATGTCCTTGT 2709
DB 181 GTAAACAAAAGCACTTAGTATTTTGGATGGCATGCCAAAACCTGTATTAATGTCCTTGT 239

RESULT 20

US-09-998-598-2127/c
Sequence 2127; Application US/09998598
Patent No. US20020150922A1
GENERAL INFORMATION:
APPLICANT: Stolk, John A.
APPLICANT: Xu, Jiangchun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO 2127
LENGTH: 230
TYPE: DNA
ORGANISM: Homo sapiens
US-09-998-598-2127

Query Match 5.3%; Score 230; DB 9; Length 230;
Best Local Similarity 100.0%; Pred. No. 5e-102;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3870 CCAGATTTAGTAGATTCATTTCTTTGTTCTCCCTCCCTGAGGACCTTTATTTTA 3929
DB 230 CCAGATTTAGTAGATTCATTTCTTTGTTCTCCCTCCCTGAGGACCTTTATTTTA 171
QY 3930 TTGTCCCTCTTCTAGATTAATTCCTCTTTGATTTGACTTTGTTAGAGAGAGTTGAC 3989
DB 170 TTGTCCCTCTTCTAGATTAATTCCTCTTTGATTTGACTTTGTTAGAGAGAGTTGAC 111
QY 3990 AGTAGATTAGCAAGTTCCAGAGTCAAAATTTAAGCTGTGTAAGCTGGGGAAATTT 4049
DB 110 AGTAGATTAGCAAGTTCCAGAGTCAAAATTTAAGCTGTGTAAGCTGGGGAAATTT 51
QY 4050 AGCTTATTTTCCCTACATGGGATCAACACTGTGTAATCAATCTTCA 4099
DB 50 AGCTTATTTTCCCTACATGGGATCAACACTGTGTAATCAATCTTCA 1

RESULT 21
US-10-242-535A-31928
Sequence 31928; Application US/10242535A
Publication No. US20040013663A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Ilew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 31928
LENGTH: 440
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (4)..(4)
OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-31928

Query Match 4.9%; Score 214; DB 17; Length 440;
Best Local Similarity 100.0%; Pred. No. 3.8e-94;
Matches 214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1240 AAAAGCCCCGTTCTTGTGAGTGACACCCCAAGAGCTCTCACTCCCAAAAAGGACCCA 1299
DB 5 AAAAGCCCCGTTCTTGTGAGTGACACCCCAAGAGCTCTCACTCCCAAAAAGGACCCA 64

QY 1300 GCAACCATCCCAAGAGAAAGCCTTCTCAAGTAGAATGAATTTGCCGTACCTTTTCCA 1359
DB 65 GCAACCATCCCAAGAGAAAGCCTTCTCAAGTAGAATGAATTTGCCGTACCTTTTCCA 124

QY 1360 CCACAGAAATGTATTTGTGTGCTTGGACACAGCCTCCCATACCGTTACATTACGGG 1419
DB 125 CCACAGAAATGTATTTGTGTGCTTGGACACAGCCTCCCATACCGTTACATTACGGG 184

QY 1420 AATCTCTCCAAAGAGAGAGAGACTGTAGCAAG 1453
DB 185 AATCTCTCCAAAGAGAGAGAGACTGTAGCAAG 218

RESULT 22
US-10-085-783A-31928
; Sequence 31928, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085, 783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31928
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4)-(4)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-31928

Query Match 4.9%; Score 214; DB 17; Length 440;
Best Local Similarity 100.0%; Pred. No. 3.8e-94;
Matches 214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1240 AAAAGCCCCGTTCTTGTGAGTGACACCCCAAGAGCTCTCACTCCCAAAAAGGACCCA 1299
DB 5 AAAAGCCCCGTTCTTGTGAGTGACACCCCAAGAGCTCTCACTCCCAAAAAGGACCCA 64

QY 1300 GCAACCATCCCAAGAGAAAGCCTTCTCAAGTAGAATGAATTTGCCGTACCTTTTCCA 1359
DB 65 GCAACCATCCCAAGAGAAAGCCTTCTCAAGTAGAATGAATTTGCCGTACCTTTTCCA 124

QY 1360 CCACAGAAATGTATTTGTGTGCTTGGACACAGCCTCCCATACCGTTACATTACGGG 1419
DB 125 CCACAGAAATGTATTTGTGTGCTTGGACACAGCCTCCCATACCGTTACATTACGGG 184

QY 1420 AATCTCTCCAAAGAGAGAGAGACTGTAGCAAG 1453
DB 185 AATCTCTCCAAAGAGAGAGAGACTGTAGCAAG 218

RESULT 23
US-09-918-995-34149

; Sequence 34149, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918, 995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235, 076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34149
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-34149

Query Match 4.9%; Score 213; DB 10; Length 411;
Best Local Similarity 99.5%; Pred. No. 1.2e-93;
Matches 383; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2179 GGCAGATTGAGAGATTTCAGGGGCTGTTCTATATACATTGGCTATGTTAAAGGGTAA 2238
DB 1 GGCAGATTGAGAGATTTCAGGGGCTGTTCTATATACATTGGCTATGTTAAAGGGTAA 60

QY 2239 AGGGCTCTCTTCTTATGACATGTGAAGTGAAGCAGCCCTTCTTTAGAGCTGTGCT 2298
DB 61 AGGGCTCTCTTCTTATGACATGTGAAGTGAAGCAGCCCTTCTTTAGAGCTGTGCT 120

QY 2299 GCATGGACCTCTTCTACCCCTGGTACACCTCTTATAGTGGTATAGATTTTAAAC 2258
DB 121 GCATGGACCTCTTCTACCCCTGGTACACCTCTTATAGTGGTATAGATTTTAAAC 180

QY 2359 CTAAATTAAGCAACCACTCAGCATG-AGCTTTAGACCAAGAGGAATGACAAGT 2417
DB 181 CTAAATTAAGCAACCACTCAGCATGAGCTTTAGACCAAGAGGAATGACAAGT 240

QY 2418 AAGCGATGAAGCAGCATCTTCACAGAGTAGAAAAAGACATCGAGAGTTGTAGTAAC 2477
DB 241 AAGCGATGAAGCAGCATCTTCACAGAGTAGAAAAAGACATCGAGAGTTGTAGTAAC 300

QY 2478 TGTCTGAAAAGATGTTGTTCAATTGAAATATTTCTGTGATACATGATGGAAAGGA 2537
DB 301 TGTCTGAAAAGATGTTGTTCAATTGAAATATTTCTGTGATACATGATGGAAAGGA 360

QY 2538 TGTGTCCTGTGATTTATTTTTCAG 2562
DB 361 TGTGTCCTGTGATTTATTTTTCAG 385

RESULT 24
US-09-736-457-1554
; Sequence 1554, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Iodes, Michael A.
; APPLICANT: Ranger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannon, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736, 457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4286 TGTCTTGAATATCTTATAAGAGTTCATTTGTAATAACTATGTGGC 4336
|||||
Db 55 TGTCTTGAATATCTTATAAGAGTTCATTTGTAATAACTATGTGGC 5
|||||

RESULT 35
US-10-191-803-451/c
; Sequence 451, Application US/10191803
; Publication No. US20040014040A1
; GENERAL INFORMATION:
; APPLICANT: MENDRICK, Donna
; APPLICANT: PORTER, Mark
; APPLICANT: JOHNSON, Kory
; APPLICANT: HIGGS, Brandon
; APPLICANT: CASTLE, Arthur
; APPLICANT: ELASHOFF, Michael
; TITLE OF INVENTION: Cardiotoxin Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5090US
; CURRENT APPLICATION NUMBER: US/10/191,803
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/303,819
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/305,623
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/369,351
; PRIOR FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: US 60/377,611
; PRIOR FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 1140
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 451
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040014040A1 AA795959
; NAME/KEY: misc_feature
; LOCATION: (1)..(590)
; OTHER INFORMATION: n = a or c or g or t
US-10-191-803-451

Query Match 1.2%; Score 51; DB 17; Length 590;
Best Local Similarity 100.0%; Pred. No. 6.4e-14;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4286 TGTCTTGAATATCTTATAAGAGTTCATTTGTAATAACTATGTGGC 4336
|||||
Db 55 TGTCTTGAATATCTTATAAGAGTTCATTTGTAATAACTATGTGGC 5
|||||

RESULT 36
US-10-106-698-3799
; Sequence 3799, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: Patentin Ver. 3.0
; SEQ ID NO 3799
; LENGTH: 196

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (7)..(8)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (39)..(39)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (70)..(70)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (76)..(76)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (78)..(78)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (128)..(128)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (143)..(143)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (146)..(146)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (180)..(180)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (196)..(196)
OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-3799

Query Match 1.1%; Score 49; DB 15; Length 196;
Best Local Similarity 100.0%; Pred. No. 6.2e-13;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3884 ATTCTATTCTTGGTTCCTCCCTCCCTGAGGACCTCTATTATTG 3932
|||||
Db 79 ATTCTATTCTTGGTTCCTCCCTCCCTGAGGACCTCTATTATTG 127
|||||

RESULT 37
US-10-425-115-167742/c
; Sequence 167742, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 167742
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_84561C.1
US-10-425-115-167742

Query Match 0.9%; Score 40; DB 18; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4333 TGGCTGTTAAAAA
|||||

```
Db      85 TGGCTGTTAAAAA 46
RESULT 38
US-10-425-115-95727/c
; Sequence 95727, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 95727
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_18817C.1
US-10-425-115-95727
Query Match      0.9%; Score 40; DB 18; Length 642;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4333 TGGCTGTTAAAAA 4372
Db      104 TGGCTGTTAAAAA 65
RESULT 39
US-10-424-599-109018
; Sequence 109018, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 109018
; LENGTH: 2627
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_69457C.1
US-10-424-599-109018
Query Match      0.9%; Score 40; DB 17; Length 2627;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4333 TGGCTGTTAAAAA 4372
Db      2455 TGGCTGTTAAAAA 2494
RESULT 40
US-10-437-963-13051/c
; Sequence 13051, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
```

```
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 13051
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_19121C.1
US-10-437-963-13051
Query Match      0.9%; Score 39; DB 18; Length 246;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4334 GGCTGTTAAAAA 4372
Db      87 GGCTGTTAAAAA 49
RESULT 41
US-10-437-963-62466/c
; Sequence 62466, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 62466
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_637C.1
US-10-437-963-62466
Query Match      0.9%; Score 39; DB 18; Length 506;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4334 GGCTGTTAAAAA 4372
Db      77 GGCTGTTAAAAA 39
RESULT 42
US-10-437-963-19880
; Sequence 19880, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
```


APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 19880
LENGTH: 2472
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_25298C.1
US-10-437-963-19880

Query Match 0.9%; Score 39; DB 18; Length 2472;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4334 GCCTGTTAAAAA 4372
DB 2430 GCCTGTTAAAAA 2468

RESULT 43
US-10-425-115-43467
Sequence 43467, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovacic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 43467
LENGTH: 376
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_139649C.1
US-10-425-115-43467

Query Match 0.9%; Score 38; DB 18; Length 376;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAA 4372
DB 80 GCTGTTAAAAA 117

RESULT 44
US-09-982-809-20
Sequence 20, Application US/09982809
Patent No. US2002010678A1
GENERAL INFORMATION:
APPLICANT: ROBISHAW, JANET
APPLICANT: KUNISCH, CHARLES
TITLE OF INVENTION: CDNA CLONES ENCODING HUMAN G PROTEIN (SUBUNITS
FILE REFERENCE: P50339
CURRENT APPLICATION NUMBER: US/09/982,809
CURRENT FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 08/952,772

PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: PCT/US95/06406
PRIOR FILING DATE: 1995-05-22
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 20
LENGTH: 827
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-982-809-20

Query Match 0.9%; Score 38; DB 9; Length 827;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAA 4372
DB 789 GCTGTTAAAAA 826

RESULT 45
US-09-925-302-156
Sequence 156, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 156
LENGTH: 1654
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-302-156

Query Match 0.9%; Score 38; DB 9; Length 1654;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAA 4372
DB 1605 GCTGTTAAAAA 1642

RESULT 46
US-09-925-302-156
Sequence 156, Application US/09925302
Publication No. US20030064072A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 156
LENGTH: 1654
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-302-156

Query Match 0.9%; Score 38; DB 10; Length 1654;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 1605 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1642

RESULT 47

US-10-425-115-141882
; Sequence 141882, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 141882
; LENGTH: 1965
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1965)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_6087C.1
US-10-425-115-141882

Query Match 0.9%; Score 38; DB 18; Length 1965;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 1589 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1626

RESULT 48

US-10-425-115-152967
; Sequence 152967, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 152967
; LENGTH: 2200
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_7108C.1
US-10-425-115-152967

Query Match 0.9%; Score 38; DB 18; Length 2200;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372

DB 2163 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2200

RESULT 49

US-10-437-963-66987
; Sequence 66987, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Bardazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 66987
; LENGTH: 2840
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_67889C.1
US-10-437-963-66987

Query Match 0.9%; Score 38; DB 18; Length 2840;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 2785 GCTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2822

RESULT 50

US-09-960-352-12974/C
; Sequence 12974, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalegan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12974
; LENGTH: 241
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 55-LIB34-056-Q1-B1-F4
US-09-960-352-12974

Query Match 0.8%; Score 37; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 107 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 71

RESULT 51
US-10-357-930-56658

```
; Sequence 56658, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-00790N
; CURRENT APPLICATION NUMBER: US/10/357,930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785,276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56658
; LENGTH: 243
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-357-930-56658
```

```
Query Match 0.8%; Score 37; DB 18; Length 243;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 155 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 191

RESULT 52
US-10-425-115-4861
; Sequence 4861, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 4861
; LENGTH: 267
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_104430C.1
US-10-425-115-4861
```

```
Query Match 0.8%; Score 37; DB 18; Length 267;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 186 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 222
```

```
RESULT 53
US-10-425-115-108211
; Sequence 108211, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 108211
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_30182C.1
US-10-425-115-108211
```

```
Query Match 0.8%; Score 37; DB 18; Length 333;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 215 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 251
```

```
RESULT 54
US-10-425-115-60416
; Sequence 60416, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 60416
; LENGTH: 361
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_1550C.1
US-10-425-115-60416
```

```
Query Match 0.8%; Score 37; DB 18; Length 361;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 240 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 276
```

```
RESULT 55
US-10-425-115-101254
; Sequence 101254, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
```

```

; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 101254
; LENGTH: 374
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MR14577_23857C.1
; US-10-425-115-101254

Query Match
Best Local Similarity 100.0%; Score 37; DB 18; Length 374;
Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAA 4372
DB 114 CTGTTAAAAA 150

RESULT 56
US-10-424-599-112320/C
; Sequence 112320, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovall, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 112320
; LENGTH: 394
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_72433C.1
; US-10-424-599-112320

Query Match
Best Local Similarity 100.0%; Score 37; DB 17; Length 394;
Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAA 4372
DB 79 CTGTTAAAAA 43

RESULT 57
US-09-918-995-8826
; Sequence 8826, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8826
```

```

; LENGTH: 410
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(410)
; OTHER INFORMATION: n = A,T,C or G
; US-09-918-995-8826

Query Match
Best Local Similarity 100.0%; Score 37; DB 10; Length 410;
Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAA 4372
DB 91 CTGTTAAAAA 127

RESULT 58
US-09-918-995-6802
; Sequence 6802, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6802
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(418)
; OTHER INFORMATION: n = A,T,C or G
; US-09-918-995-6802

Query Match
Best Local Similarity 100.0%; Score 37; DB 10; Length 418;
Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAA 4372
DB 57 CTGTTAAAAA 93

RESULT 59
US-10-425-115-142294/C
; Sequence 142294, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovall, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 142294
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MR14577_61252C.1
```

US-10-425-115-142294

Query Match 0.8%; Score 37; DB 18; Length 421;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4336 CTGTTAAAAA
Db 120 CTGTTAAAAA
4372
84

RESULT 60

US-10-424-599-3745/c
; Sequence 3745, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 3745
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_103388C.1
US-10-424-599-3745

Query Match 0.8%; Score 37; DB 17; Length 424;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4336 CTGTTAAAAA
Db 84 CTGTTAAAAA
4372
48

RESULT 61

US-09-918-995-10500
; Sequence 10500, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10500
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(450)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-10500

Query Match 0.8%; Score 37; DB 10; Length 450;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4336 CTGTTAAAAA
4372

Db 339 CTGTTAAAAA
375

RESULT 62

US-10-425-115-51597/c
; Sequence 51597, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 51597
; LENGTH: 464
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_147050C.1
US-10-425-115-51597

Query Match 0.8%; Score 37; DB 18; Length 464;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4336 CTGTTAAAAA
Db 80 CTGTTAAAAA
4372
44

RESULT 63

US-10-425-115-50595
; Sequence 50595, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 50595
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_146138C.1
US-10-425-115-50595

Query Match 0.8%; Score 37; DB 18; Length 583;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4336 CTGTTAAAAA
Db 266 CTGTTAAAAA
4372
302

RESULT 64

US-09-992-331-4
; Sequence 4, Application US/09992331
; Publication No. US20030022186A1
; GENERAL INFORMATION:

```

; APPLICANT: FEDER, JOHN N.
; APPLICANT: MINTER, GABE
; APPLICANT: RAMANATHAN, CHANDRA S.
; APPLICANT: HAWKEN, DONALD R.
; TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPBMY18,
; TITLE OF INVENTION: EXPRESSED HIGHLY IN PITUITARY GLAND AND COLON CARCINOMA
; FILE REFERENCE: D0048NP
; CURRENT APPLICATION NUMBER: US/09/992,331
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/308,540
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/261,782
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 60/248,483
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 632
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-992-331-4

Query Match
Best Local Similarity 100.0%; Score 37; DB 10; Length 632;
Pred. No. 5e-07; Mismatches 0; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 582 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 618
```

```

RESULT 65
; Sequence 4, Application US/10262313
; Publication No. US20030129653A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPBMY18, EXPRESSED HIGH
; FILE REFERENCE: D0048 CIP
; CURRENT APPLICATION NUMBER: US/10/262,313
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: U.S. 09/992,331
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 632
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-262-313-4

Query Match
Best Local Similarity 100.0%; Score 37; DB 15; Length 632;
Pred. No. 5e-07; Mismatches 0; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 582 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 618
```

```

RESULT 66
; Sequence 4, Application US/10768878
; Publication No. US20040161823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPBMY18, EXPRESSED
; FILE REFERENCE: D0048A CIP2
; CURRENT APPLICATION NUMBER: US/10/768,878
```

```

; CURRENT FILING DATE: 2004-01-30
; PRIOR APPLICATION NUMBER: U.S. 09/992,331
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: U.S. 60/248,483
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: U.S. 10/262,313
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: U.S. 60/261,782
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: U.S. 60/308,540
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 632
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-768-878-4

Query Match
Best Local Similarity 100.0%; Score 37; DB 18; Length 632;
Pred. No. 5e-07; Mismatches 0; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 582 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 618
```

```

RESULT 67
; Sequence 66141, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Zhou, Yihua
; APPLICANT: Zhoul, David K
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 66141
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3073C.1
; US-10-424-599-66141

Query Match
Best Local Similarity 100.0%; Score 37; DB 17; Length 642;
Pred. No. 5e-07; Mismatches 0; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 67 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 31
```

```

RESULT 68
; Sequence 1661, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovall, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
```

;; CURRENT FILING DATE: 2003-04-28
;; NUMBER OF SEQ ID NOS: 369326
;; SEQ ID NO 1661
;; LENGTH: 643
;; TYPE: DNA
;; ORGANISM: Zea mays
;; FEATURE:
;; OTHER INFORMATION: Clone ID: MRT4577_10151C.1
US-10-425-115-1661

Query Match 0.8%; Score 37; DB 18; Length 643;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAA
DB 95 CTGTTAAAAA
59

RESULT 69
US-10-313-542-54/C
; Sequence 54, Application US/10313542
; Publication No. US20030120057A1
; GENERAL INFORMATION:
; APPLICANT: Roopa, Reddy
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED F
; FILE REFERENCE: PA-0013 US
; CURRENT APPLICATION NUMBER: US/10/313,542
; CURRENT FILING DATE: 2002-12-05
; PRIOR APPLICATION NUMBER: US/09/495,050
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118,318
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: PERL Program
; SEQ ID NO 54
; LENGTH: 647
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030120057A1 1288503CT1
US-10-313-542-54

Query Match 0.8%; Score 37; DB 15; Length 647;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAA
DB 83 CTGTTAAAAA
47

RESULT 70
US-10-425-115-15766/C
; Sequence 15766, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 15766
; LENGTH: 648
; TYPE: DNA
; ORGANISM: Homo sapiens

;; ORGANISM: Zea mays
;; FEATURE:
;; OTHER INFORMATION: Clone ID: MRT4577_114378C.1
US-10-425-115-15766

Query Match 0.8%; Score 37; DB 18; Length 648;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAA
DB 113 CTGTTAAAAA
77

RESULT 71
US-10-437-963-37096/C
; Sequence 37096, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 37096
; LENGTH: 661
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_40857C.1
US-10-437-963-37096

Query Match 0.8%; Score 37; DB 18; Length 661;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAA
DB 91 CTGTTAAAAA
55

RESULT 72
US-10-198-846-8291/C
; Sequence 8291, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8291
; LENGTH: 674
; TYPE: DNA
; ORGANISM: Homo sapiens

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4, 6, 237, 258, 266, 288, 289, 293, 295, 302, 318, 327, 337,
; LOCATION: 361, 366, 367, 376, 381, 382, 384, 389, 390, 396, 399, 400,
; LOCATION: 404, 405, 411, 417, 423, 426, 428, 430, 433, 435, 438,
; LOCATION: 447, 448, 449, 451, 452, 455, 457, 462, 463, 466
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 468, 469, 471, 474, 480, 481, 489, 490, 491, 493, 497, 498,
; LOCATION: 510, 518, 520, 522, 523, 525, 526, 527, 537, 538, 540, 545,
; LOCATION: 546, 549, 550, 553, 554, 562, 565, 570, 571, 572, 575, 576,
; LOCATION: 578, 581, 583, 587, 590, 591, 594, 595, 600, 602, 605
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 607, 608, 609, 611, 614, 616, 618, 620, 628, 632, 636, 637,
; LOCATION: 638, 647, 650, 653, 654, 662, 664, 669, 672, 672, 674
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-8291
```

```

Query Match      0.8%; Score 37; DB 14; Length 674;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      89 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 53
```

```

RESULT 73
US-10-437-963-32884
; Sequence 32884, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: la Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 32884
; LENGTH: 698
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(698)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37050C.1
US-10-437-963-32884
```

```

Query Match      0.8%; Score 37; DB 18; Length 698;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      599 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 635
```

```

RESULT 74
US-10-425-115-164365
; Sequence 164365, Application US/10425115
```

```

; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: la Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 164365
; LENGTH: 722
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_81477C.1
US-10-425-115-164365
```

```

Query Match      0.8%; Score 37; DB 18; Length 722;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      663 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 699
```

```

RESULT 75
US-10-106-698-316
; Sequence 316, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 316
; LENGTH: 782
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-106-698-316
```

```

Query Match      0.8%; Score 37; DB 15; Length 782;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      729 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 765
```

```

Search completed: March 25, 2005, 19:49:49
Job time : 2747 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2005, 05:02:59 ; Search time 672 Seconds
(without alignments)
10645.538 Million cell updates/sec

Title: US-10-054-935-1

Perfect score: 4372

Sequence: 1 cagctcctgcagcccccgcac.....aaaaaaaaaaaaaa 4372

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1202784 seqs, 818138359 residues

Word size : 15

Total number of hits satisfying chosen parameters: 76027

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/5A COMB.seq:*
- 2: /cgn2_6/prodata/1/ina/5B COMB.seq:*
- 3: /cgn2_6/prodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/prodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/prodata/1/ina/6C COMB.seq:*
- 6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	903	20.7	1545	4 US-09-023-655-167	Sequence 167, App
2	487	11.1	1230	4 US-09-023-655-122	Sequence 322, App
3	152	3.5	542	4 US-09-702-705-1554	Sequence 1554, App
4	152	3.5	542	4 US-09-736-457-1554	Sequence 1554, App
5	152	3.5	542	4 US-09-614-124B-1554	Sequence 1554, App
6	152	3.5	542	4 US-09-671-325-1554	Sequence 1554, App
7	152	3.5	542	4 US-09-658-824-1554	Sequence 1554, App
8	106	2.4	317	4 US-09-513-999C-29348	Sequence 29348, A
9	47	1.1	50	4 US-09-513-999C-15253	Sequence 15253, A
10	38	0.9	138	4 US-09-621-976-9595	Sequence 9595, App
11	38	0.9	827	5 PCT-US95-064064-20	Sequence 20, App1
12	37	0.8	100	4 US-09-621-976-12774	Sequence 12774, A
13	37	0.8	458	4 US-08-524-757-1	Sequence 1, App1
14	37	0.8	647	4 US-09-495-050A-54	Sequence 54, App1
15	37	0.8	1706	4 US-10-033-109-3	Sequence 3, App1
16	37	0.8	1883	3 US-09-149-476-170	Sequence 170, App
17	37	0.8	2044	4 US-09-949-016-646	Sequence 646, App
18	37	0.8	3366	4 US-09-556-141C-6	Sequence 6, App1
19	37	0.8	3366	4 US-09-555-526C-6	Sequence 6, App1
20	37	0.8	10442	4 US-09-596-141C-1	Sequence 1, App1
21	37	0.8	10442	4 US-09-595-526C-1	Sequence 1, App1
22	37	0.8	10474	4 US-09-596-141C-7	Sequence 7, App1
23	37	0.8	10474	4 US-09-596-141C-9	Sequence 9, App1
24	37	0.8	10474	4 US-09-595-526C-7	Sequence 7, App1
25	37	0.8	10474	4 US-09-595-526C-9	Sequence 9, App1
26	36	0.8	120	1 US-08-153-051B-28	Sequence 28, App1
27	36	0.8	120	1 US-08-060-952C-44	Sequence 44, App1

28	36	0.8	120	2 US-08-151-477A-28	Sequence 28, App1
29	36	0.8	120	3 US-08-819-867-58	Sequence 58, App1
30	36	0.8	120	3 US-08-464-011B-44	Sequence 44, App1
31	36	0.8	120	4 US-09-378-535-58	Sequence 58, App1
32	36	0.8	168	4 US-09-621-976-15316	Sequence 15316, A
33	36	0.8	173	4 US-09-621-976-15319	Sequence 15319, A
34	36	0.8	178	4 US-09-621-976-15322	Sequence 15322, A
35	36	0.8	180	4 US-09-621-976-15320	Sequence 15320, A
36	36	0.8	194	4 US-09-621-976-15317	Sequence 15317, A
37	36	0.8	223	3 US-09-792-594-11	Sequence 11, App1
38	36	0.8	284	4 US-09-621-976-2841	Sequence 2841, App
39	36	0.8	601	4 US-09-949-016-185724	Sequence 185724, App
40	36	0.8	675	4 US-09-800-729-21	Sequence 21, App1
41	36	0.8	675	4 US-09-621-976-2461	Sequence 2461, App
42	36	0.8	997	4 US-09-907-994A-376	Sequence 376, App
43	36	0.8	997	4 US-09-905-125A-376	Sequence 376, App
44	36	0.8	997	4 US-09-902-775A-376	Sequence 376, App
45	36	0.8	997	4 US-09-906-700-376	Sequence 376, App
46	36	0.8	997	4 US-09-903-603A-376	Sequence 376, App
47	36	0.8	997	4 US-09-904-920A-376	Sequence 376, App
48	36	0.8	997	4 US-09-909-064-376	Sequence 376, App
49	36	0.8	997	4 US-09-905-381A-376	Sequence 376, App
50	36	0.8	997	4 US-09-906-618-376	Sequence 376, App
51	36	0.8	1094	4 US-09-522-714-1	Sequence 1, App1
52	36	0.8	1123	3 US-09-152-060-15	Sequence 15, App1
53	36	0.8	1281	4 US-09-436-521A-5	Sequence 5, App1
54	36	0.8	1781	3 US-09-499-302A-1	Sequence 1, App1
55	36	0.8	7724	4 US-08-486-049-1	Sequence 1, App1
56	36	0.8	28686	4 US-09-949-016-17054	Sequence 17054, A
57	36	0.8	115	4 US-09-949-016-12335	Sequence 12335, A
58	35	0.8	187	4 US-09-621-976-13418	Sequence 13418, A
59	35	0.8	350	1 US-08-248-016-11	Sequence 11, App1
60	35	0.8	350	1 US-08-451-501-11	Sequence 11, App1
61	35	0.8	350	5 PCT-US95-06761-11	Sequence 11, App1
62	35	0.8	413	4 US-09-621-976-16712	Sequence 16712, A
63	35	0.8	425	4 US-09-621-976-16710	Sequence 16710, A
64	35	0.8	430	4 US-09-621-976-16711	Sequence 16711, A
65	35	0.8	430	4 US-09-621-976-2632	Sequence 2632, App
66	35	0.8	445	4 US-09-621-976-2630	Sequence 2630, App
67	35	0.8	794	3 US-09-385-982-537	Sequence 537, App
68	35	0.8	801	4 US-10-039-659A-5	Sequence 5, App1
69	35	0.8	1192	3 US-09-318-448-18	Sequence 18, App1
70	35	0.8	1355	4 US-09-614-221A-3	Sequence 3, App1
71	35	0.8	1414	4 US-09-501-115-5	Sequence 5, App1
72	35	0.8	1508	4 US-09-039-046-1	Sequence 1, App1
73	35	0.8	1515	4 US-09-205-258-40	Sequence 40, App1
74	35	0.8	1604	1 US-08-665-966-9	Sequence 9, App1
75	35	0.8	1604	3 US-09-041-780-9	Sequence 9, App1
76	35	0.8	1748	4 US-08-635-967-1	Sequence 1, App1
77	35	0.8	1799	3 US-08-687-590-58	Sequence 58, App1
78	35	0.8	1825	4 US-09-187-899-8	Sequence 8, App1
79	35	0.8	1868	1 US-08-658-883B-1	Sequence 1, App1
80	35	0.8	1868	3 US-09-676-610B-26	Sequence 26, App1
81	35	0.8	2075	4 US-09-461-325-130	Sequence 130, App
82	35	0.8	2075	4 US-10-012-542-130	Sequence 130, App
83	35	0.8	2075	4 US-10-115-123-130	Sequence 130, App
84	35	0.8	2075	4 US-09-541-782-7	Sequence 7, App1
85	35	0.8	3709	3 US-09-723-820-7	Sequence 7, App1
86	35	0.8	3709	4 US-10-270-085-7	Sequence 7, App1
87	35	0.8	4628	4 US-09-578-063-9	Sequence 9, App1
88	35	0.8	6412	4 US-09-769-987-1	Sequence 1, App1
89	35	0.8	63760	4 US-09-949-016-14087	Sequence 14087, A
90	35	0.8	63760	4 US-09-949-016-14088	Sequence 14088, A
91	35	0.8	183202	4 US-09-949-016-13614	Sequence 13614, A
92	34	0.8	60	3 US-09-284-627-14	Sequence 14, App1
93	34	0.8	80	3 US-09-284-627-15	Sequence 15, App1
94	34	0.8	85	4 US-09-621-976-14741	Sequence 14741, A
95	34	0.8	105	3 US-09-284-627-23	Sequence 23, App1
96	34	0.8	114	1 US-08-120-827-99	Sequence 99, App1
97	34	0.8	114	1 US-08-478-675-99	Sequence 99, App1
98	34	0.8	141	4 US-09-621-976-8705	Sequence 8705, App
99	34	0.8	144	1 US-08-702-344-26	Sequence 26, App1
100	34	0.8	144	1 US-08-702-344-26	Sequence 26, App1

ALIGNMENTS

```
RESULT 1
US-09-023-655-167
/ Sequence 167, Application US/09023655
/ Patent No. 6607879
/ GENERAL INFORMATION:
/ APPLICANT: Cocks, Benjamin G.
/ APPLICANT: Susan G. Stuart
/ APPLICANT: Jeffrey J. Sellhammer
/ TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
/ TITLE OF INVENTION: EXPRESSION
/ NUMBER OF SEQUENCES: 1508
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
/ STREET: 3174 PORTER DRIVE
/ CITY: PALO ALTO
/ STATE: CALIFORNIA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/023,655
/ FILING DATE: HERewith
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Zeller, Karen J.
/ REGISTRATION NUMBER: 37,071
/ REFERENCE/DOCKET NUMBER: PA-0001 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650) 855-0555
/ TELEFAX: (650) 845-4166
/ INFORMATION FOR SEQ ID NO: 167:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1545 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: TBLYN0701
/ CLONE: 040476
/ US-09-023-655-167

Query Match      20.7%; Score 903; DB 4; Length 1545;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 3666 GTGTCCAAAAAGAAAAAGCCCAATCACTGTCTCTTGAAGTTGTTCTTTGATCCCTCAG 3725
DB 871 GTGTCCAAAAAGAAAAAGCCCAATCACTGTCTCTTGAAGTTGTTCTTTGATCCCTCAG 930
QY 3726 TTTCTCTGTAATTTGAGCATGTGTGGGTTCCAAATTTTGGGATAGATTAGCAAAATTTA 3785
DB 931 TTTCTCTGTAATTTGAGCATGTGTGGGTTCCAAATTTTGGGATAGATTAGCAAAATTTA 990
QY 3786 ACCATTGTGTTGTGCTTACCCAGAGGAGCTCCCAAGTTTGTGACTTGAAGTAGAGTAG 3845
DB 991 ACCATTGTGTTGTGCTTACCCAGAGGAGCTCCCAAGTTTGTGACTTGAAGTAGAGTAGAG 1050
QY 3846 AAGAATCCAGAGGTCTATCTGGCCAGATTAAAGTAGATTCTATTTCTGTGTTCTCCC 3905
DB 1051 AAGAATCCAGAGGTCTATCTGGCCAGATTAAAGTAGATTCTATTTCTGTGTTCTCCC 1110
QY 3906 TCTCCCTGAGGACCTCTTATTTTATTTGTCCTCCCTCTGAGTTAATCTCTCTGATTTG 3965
DB 1111 TCTCCCTGAGGACCTCTTATTTTATTTGTCCTCCCTCTGAGTTAATCTCTCTGATTTG 1170
QY 3966 ACTTTGTGAGAGGAGGTTGACAGTAGATTAGCAAAAGTTCCAAAGTGCAAAATTAAGT 4025
DB 1171 ACTTTGTGAGAGGAGGTTGACAGTAGATTAGCAAAAGTTCCAAAGTGCAAAATTAAGT 1230
QY 4026 GTGTGAGGTGTGGGGGAAAATTAGTTATTTTCTTCTTCACTGGGATCAACTGTG 4085
DB 1231 GTGTGAGGTGTGGGGGAAAATTAGTTATTTTCTTCTTCACTGGGATCAACTGTG 1290
QY 4086 AATTCAATCTTCACTGAAAGGCCCTGACGTTCCCTAAACAGTTGTTGTTTCTT 4145
DB 1291 AATTCAATCTTCACTGAAAGGCCCTGACGTTCCCTAAACAGTTGTTGTTTCTT 1350
QY 4146 TAACAAAGTTTAAAGTAGTTAATTAATTAATAAAATTTGCTGTCTGCTCACTCAG 4205
DB 1351 TAACAAAGTTTAAAGTAGTTAATTAATTAATAAAATTTGCTGTCTGCTCACTCAG 1410
QY 4206 TTTGTTTATGCCCATTTTCATATTTGTGTCTGTGTGTAATTCATAACTTTGATACAT 4265
DB 1411 TTTGTTTATGCCCATTTTCATATTTGTGTCTGTGTGTAATTCATAACTTTGATACAT 1470
QY 4266 TTTGATGCTTAATAATTTGTGTCTGTGTAATTTCTTAATAAGTCAATCTTAATA 4325
DB 1471 TTTGATGCTTAATAATTTGTGTCTGTGTAATTTCTTAATAAGTCAATCTTAATA 1530
QY 4326 ACT 4328
DB 1531 ACT 1533

RESULT 2
US-09-023-655-322
/ Sequence 322, Application US/09023655
/ Patent No. 6607879
/ GENERAL INFORMATION:
/ APPLICANT: Cocks, Benjamin G.
/ APPLICANT: Susan G. Stuart
/ APPLICANT: Jeffrey J. Sellhammer
/ TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
/ TITLE OF INVENTION: EXPRESSION
/ NUMBER OF SEQUENCES: 1508
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
/ STREET: 3174 PORTER DRIVE
/ CITY: PALO ALTO
/ STATE: CALIFORNIA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
```


Db	66	TGTGTGNTGTGTGTGTGTGTGCTAATGGGTTTTTCATTTGTAACTCCATTCGCTTAAGAG	125
QY	3455	AGTGGGCTCTCTATPAAGGGAACCGGCTGTAACCTTACGACGAAGAATGPAAGAGAA	3514
Db	126	AGTGGGCTCTCTATPAAGGGAACCGGCTGTAACCTTACGACGAAGAATGPAAGAGAA	185
QY	3515	ATAGGACTTAAATCCACTAGGGGGCTCTCATCTCACACCTTAAAGAGAGATTTCTAGAA	3574
Db	186	ATAGGACTTAAATCCACTAGGGGGCTCTCATCTCACACCTTAAAGAGAGATTTCTAGAA	245
QY	3575	AACGGGCGCAATTTCTTGTGTCGCATCAATTTAAATGTGGCAGGCGTTACAGTTCT	3634
Db	246	AACGGGCGCAATTTCTTGTGTCGCATCAATTTAAATGTGGCAGGCGTTACAGTTCT	305
QY	3635	TACTCTTAACGATGTGATATTTCTTCGTAACTGTCCAAAAAGAAAAAGAACCCATCAG	3694
Db	306	TACTCTTAACGATGTGATATTTCTTCGTAACTGTCCAAAAAGAAAAAGAACCCATCAG	365
QY	3695	TGTGCTTGAATTTGTTCTTTGATGCCCAAGTTTCTCTGATATTTACAGATGTGC	3750
Db	366	TGTGCTTGAATTTGTTCTTTGATGCCCAAGTTTCTCTGATATTTACAGATGTGC	421

RESULT 7
US-09-658-824-1554
; Sequence 1554, Application US/09658824

GENERAL INFORMATION:

```

APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darriick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C11
CURRENT APPLICATION NUMBER: US/09/658,824
CURRENT FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 1788
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 1554
LENGTH: 542
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)_(542)
OTHER INFORMATION: n = A,T,C or G
US-09-658-824-1554

```

Query Match . . . 3.5%; Score 152; DB 4; Length 542;

Best Local Similarity 98.9%; Pred. No. 1.8e-53;
Matches 352; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

[illegible][illegible]

RESULT 8
US-09-513-999C-29348

; Sequence 29348, Application US/095139999C
; Patent No. 6783961

```

; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.

```

! APPLICANT: Duclert, A.
! APPLICANT: Giordano, J.Y.

1 ; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
2 ; Patent No. 6783961

```

; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C

```

;
CURRENT FILING DATE: 2000-02-24
;
PRIOR APPLICATION NUMBER: US 60/122,487
;

; PRIOR FILING DATE: 1999-02-
 ; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Pate
; SEQ ID NO 29348

```

; LENGTH: 317
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:

```

```

; NAME/KEY: misc_feature
;. LOCATION: 13

```

```

;
OTHER INFORMATION: h=a
FEATURE:

```

```
; NAME/KEY: misc_feature
; LOCATION: 17
```

```

; OTHER INFORMATION: r=a or g
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: 50

```

```

; OTHER INFORMATION: r=a or g
;
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: 51
;

```

```

; OTHER INFORMATION: w=a or t
; FEATURE:
;

```

```

; NAME/KEY: misc_feature
; LOCATION: 52

```

```

; OTHER INFORMATION: r=a or g
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: 53

```

```

; OTHER INFORMATION: w=a or b
; FEATURE:

```

```
; NAME/KEY: misc_feature
; LOCATION: 145
```

```

; OTHER INFORMATION: S=g or c
; FEATURE:

```

```
; NAME/KEY: misc_feature
; LOCATION: 146
```

```

; OTHER INFORMATION: S=g or c
; FEATURE:
;

```

```

; NAME/KEY: misc_feature
; LOCATION: 158
;

```

OTHER INFORMATION;
US-09-513-999C-29348

Query Match 2.4%; Score 106; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. NO. 2.7e-34;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3569 TAGAAAACTGGCCAGATTTCTTGTCTCCATCTTTAAATGTCGACGGCTTACG 3628
DB 212 TAGAAAACTGGCCAGATTTCTTGTCTCCATCTTTAAATGTCGACGGCTTACG 271
QY 3629 TTTTCTACTCTTACCTATGATATTTCTTCTGTAACGTCCAAA 3674
DB 272 TTTTCTACTCTTACCTATGATATTTCTTCTGTAACGTCCAAA 317

RESULT 9
US-09-513-999C-15253
; Sequence 15253, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 15253
; LENGTH: 50
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-15253

Query Match 1.1%; Score 47; DB 4; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 AAACACCTCACCGAGCTGTAGAAATATGCTGTGCGCAAGAACCC 1940
DB 1 AAACACCTCACCGAGCTGTAGAAATATGCTGTGCGCAAGAACCC 47
RESULT 10
US-09-621-976-9595
; Sequence 9595, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 9595
; LENGTH: 138
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-9595

Query Match 0.9%; Score 38; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 6.1e-06;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAA 4372
DB 68 GCTGTTAAAAA 105

RESULT 11
PCT-US95-06406A-20
; Sequence 20, Application PC/TUS9506406A

GENERAL INFORMATION:
; APPLICANT: Janet D. Robshaw, Charles Kunsch
; TITLE OF INVENTION: cDNA Clones Encoding Human G Protein
; TITLE OF INVENTION: Subunits
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06406A
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 827
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: No
PCT-US95-06406A-20

Query Match 0.9%; Score 38; DB 5; Length 827;
Best Local Similarity 100.0%; Pred. No. 5.7e-06;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4335 GCTGTTAAAAA 4372
DB 789 GCTGTTAAAAA 826
RESULT 12
US-09-621-976-12774
; Sequence 12774, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 12774
; LENGTH: 100
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-12774

Query Match 0.8%; Score 37; DB 4; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 25 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 61

RESULT 13
US-08-524-757-1
; Sequence 1, Application US/08524757
; Patent No. 5792634
; GENERAL INFORMATION:
; APPLICANT: Conaway, Ronald C.
; APPLICANT: Conaway, Joan W.
; APPLICANT: Bradsher, John N.
; TITLE OF INVENTION: RNA Polymerase Transcription Factor
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: TX
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/524,757
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13621
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/160087
; FILING DATE: 30-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Harre, John A.
; REGISTRATION NUMBER: 37,345
; REFERENCE/DOCKET NUMBER: B35006CIPCIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (214) 939-4500
; TELEFAX: (214) 939-4600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 46..381
US-08-524-757-1

Query Match 0.8%; Score 37; DB 1; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 405 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 441

RESULT 14
US-09-495-050A-54/c
; Sequence 54, Application US/09495050A
; Patent No. 6492505
; GENERAL INFORMATION:
; APPLICANT: Roopa, Reddy
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice

; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P
; FILE REFERENCE: PA-0013 US
; CURRENT APPLICATION NUMBER: US/09/495,050A
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118,318
; PRIOR FILING DATE: February 1, 1999
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: PERL Program
; SEQ ID NO 54
; LENGTH: 647
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6492505 1288503CT1
US-09-495-050A-54

Query Match 0.8%; Score 37; DB 4; Length 647;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 83 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 47

RESULT 15
US-10-033-109-3
; Sequence 3, Application US/10033109
; Patent No. 6833492
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Rafaleki, J. Antoni
; APPLICANT: Sakai, Hajime
; TITLE OF INVENTION: Nitrogen Transport Metabolism
; FILE REFERENCE: BB-1210
; CURRENT APPLICATION NUMBER: US/10/033,109
; CURRENT FILING DATE: 2001-12-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/384,625
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/098,248
; PRIOR FILING DATE: EARLIER FILING DATE: 28 August 1998
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
; LENGTH: 1706
; TYPE: DNA
; ORGANISM: Glycine max
US-10-033-109-3

Query Match 0.8%; Score 37; DB 4; Length 1706;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1670 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1706

RESULT 16
US-09-149-476-170
; Sequence 170, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07

[illegible]


```

; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match          0.8%; Score 37; DB 3; Length 1883;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      1841 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1877

RESULT 17
US-09-949-016-646
; Sequence 646, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 646
; LENGTH: 2044
```

```

; TYPE: DNA
; ORGANISM: Human
US-09-949-016-646

Query Match          0.8%; Score 37; DB 4; Length 2044;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      2006 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2042

RESULT 18
US-09-596-141C-6
; Sequence 6, Application US/09596141C
; Patent No. 6821774
; GENERAL INFORMATION:
; APPLICANT: Lawn, Richard M.
; APPLICANT: Wade, David
; APPLICANT: Oram, John F.
; APPLICANT: Garvin, Michael
; TITLE OF INVENTION: Compositions and Methods for Increasing Cholesterol
; TITLE OF INVENTION: Efflux and Raising HDL using ATP Binding Cassette
; FILE REFERENCE: 99,395-B
; CURRENT APPLICATION NUMBER: US/09/596,141C
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 60/140,264
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 60/153,872
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 60/166,573
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 62
; SEQ ID NO 6
; LENGTH: 3366
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(3366)
; OTHER INFORMATION: All n's are unknown.
US-09-596-141C-6

Query Match          0.8%; Score 37; DB 4; Length 3366;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      3286 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3322

RESULT 19
US-09-595-526C-6
; Sequence 6, Application US/09595526C
; Patent No. 6835563
; GENERAL INFORMATION:
; APPLICANT: Lawn, Richard M.
; APPLICANT: Wade, David
; APPLICANT: Oram, John F.
; APPLICANT: Garvin, Michael
; TITLE OF INVENTION: Compositions and Methods for Increasing Cholesterol
; TITLE OF INVENTION: Efflux and Raising HDL using ATP Binding Cassette
; FILE REFERENCE: 99,395-A
; CURRENT APPLICATION NUMBER: US/09/595,526C
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 60/140,264
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 60/153,872
; PRIOR FILING DATE: 1999-09-14
```

```

; PRIOR APPLICATION NUMBER: US 60/166,573
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 62
; SEQ ID NO 6
; LENGTH: 3366
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(3366)
; OTHER INFORMATION: All n's are unknown.
US-09-595-526C-6
```

```

Query Match          0.8%; Score 37; DB 4; Length 3366;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB      3286 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3322
```

RESULT 20

```

US-09-596-141C-1
; Sequence 1, Application US/09596141C
; Patent No. 6821774
; GENERAL INFORMATION:
; APPLICANT: Lawn, Richard M.
; APPLICANT: Wade, David
; APPLICANT: Oram, John F.
; APPLICANT: Garvin, Michael
; TITLE OF INVENTION: Compositions and Methods for Increasing Cholesterol
; TITLE OF INVENTION: Efflux and Raising HDL using ATP Binding Cassette
; FILE REFERENCE: 99,395-B
; CURRENT APPLICATION NUMBER: US/09/596,141C
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 60/140,264
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 60/153,872
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 60/166,573
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 62
; SEQ ID NO 1
; LENGTH: 10442
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(10442)
; OTHER INFORMATION: All n's are unknown.
US-09-596-141C-1
```

```

Query Match          0.8%; Score 37; DB 4; Length 10442;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB      10362 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 10398
```

RESULT 21

```

US-09-595-526C-1
; Sequence 1, Application US/09595526C
; Patent No. 6835563
; GENERAL INFORMATION:
; APPLICANT: Lawn, Richard M.
; APPLICANT: Wade, David
; APPLICANT: Oram, John F.
; APPLICANT: Garvin, Michael
; TITLE OF INVENTION: Compositions and Methods for Increasing Cholesterol
```

```

; TITLE OF INVENTION: Efflux and Raising HDL using ATP Binding Cassette
; FILE REFERENCE: 99,395-A
; CURRENT APPLICATION NUMBER: US/09/595,526C
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 60/140,264
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 60/153,872
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 60/166,573
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 62
; SEQ ID NO 1
; LENGTH: 10442
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(10442)
; OTHER INFORMATION: All n's are unknown.
US-09-595-526C-1
```

```

Query Match          0.8%; Score 37; DB 4; Length 10442;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB      10362 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 10398
```

RESULT 22

```

US-09-596-141C-7
; Sequence 7, Application US/09596141C
; Patent No. 6821774
; GENERAL INFORMATION:
; APPLICANT: Lawn, Richard M.
; APPLICANT: Wade, David
; APPLICANT: Oram, John F.
; APPLICANT: Garvin, Michael
; TITLE OF INVENTION: Compositions and Methods for Increasing Cholesterol
; TITLE OF INVENTION: Efflux and Raising HDL using ATP Binding Cassette
; FILE REFERENCE: 99,395-B
; CURRENT APPLICATION NUMBER: US/09/596,141C
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 60/140,264
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 60/153,872
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 60/166,573
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 62
; SEQ ID NO 7
; LENGTH: 10474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(10474)
; OTHER INFORMATION: All n's are unknown.
US-09-596-141C-7
```

```

Query Match          0.8%; Score 37; DB 4; Length 10474;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB      10394 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 10430
```

RESULT 23

US-09-596-141C-9
; Sequence 9, Application US/09596141C
; Patent No. 6821774
; GENERAL INFORMATION:
; APPLICANT: Lawn, Richard M.
; APPLICANT: Wade, David
; APPLICANT: Oram, John F.
; APPLICANT: Garvin, Michael
; TITLE OF INVENTION: Compositions and Methods for Increasing Cholesterol
; TITLE OF INVENTION: Efflux and Raising HDL using ATP Binding Cassette
; TITLE OF INVENTION: Transporter Protein ABC1
; FILE REFERENCE: 99,395-B
; CURRENT APPLICATION NUMBER: US/09/596,141C
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 60/140,264
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 60/153,872
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 60/166,573
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 62
; SEQ ID NO 9
; LENGTH: 10474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(10474)
; OTHER INFORMATION: All n's are unknown.
US-09-596-141C-9

Query Match 0.8%; Score 37; DB 4; Length 10474;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 10394 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 10430

RESULT 24
US-09-595-526C-7
; Sequence 7, Application US/09595526C
; Patent No. 6835563
; GENERAL INFORMATION:
; APPLICANT: Lawn, Richard M.
; APPLICANT: Wade, David
; APPLICANT: Oram, John F.
; APPLICANT: Garvin, Michael
; TITLE OF INVENTION: Compositions and Methods for Increasing Cholesterol
; TITLE OF INVENTION: Efflux and Raising HDL using ATP Binding Cassette
; TITLE OF INVENTION: Transporter Protein ABC1
; FILE REFERENCE: 99,395-A
; CURRENT APPLICATION NUMBER: US/09/595,526C
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 60/140,264
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 60/153,872
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 60/166,573
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 62
; SEQ ID NO 7
; LENGTH: 10474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(10474)
; OTHER INFORMATION: All n's are unknown.
US-09-595-526C-7

Query Match 0.8%; Score 37; DB 4; Length 10474;

Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 10394 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 10430

RESULT 25
US-09-595-526C-9
; Sequence 9, Application US/09595526C
; Patent No. 6835563
; GENERAL INFORMATION:
; APPLICANT: Lawn, Richard M.
; APPLICANT: Wade, David
; APPLICANT: Oram, John F.
; APPLICANT: Garvin, Michael
; TITLE OF INVENTION: Compositions and Methods for Increasing Cholesterol
; TITLE OF INVENTION: Efflux and Raising HDL using ATP Binding Cassette
; TITLE OF INVENTION: Transporter Protein ABC1
; FILE REFERENCE: 99,395-A
; CURRENT APPLICATION NUMBER: US/09/595,526C
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 60/140,264
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 60/153,872
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 60/166,573
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 62
; SEQ ID NO 9
; LENGTH: 10474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(10474)
; OTHER INFORMATION: All n's are unknown.
US-09-595-526C-9

Query Match 0.8%; Score 37; DB 4; Length 10474;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 10394 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 10430

RESULT 26
US-08-153-051B-28
; Sequence 28, Application US/08153051B
; Patent No. 5645986
; GENERAL INFORMATION:
; APPLICANT: Michael D. West
; APPLICANT: Jerry W. Shay
; APPLICANT: Woodring E. Wright
; APPLICANT: Elizabeth Blackburn
; APPLICANT: Nam Woo Kim
; APPLICANT: Calvin B. Harley
; APPLICANT: Scott L. Weinrich
; APPLICANT: Catherine Strahl
; APPLICANT: Michael J. McEachern
; APPLICANT: Homayoun Vaziri
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: CONDITIONS RELATED TO TELOMERE
; TITLE OF INVENTION: LENGTH AND/OR TELOMERASE ACTIVITY
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles

STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,051B
FILING DATE: No. 5645986ember 12, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/038,766
FILING DATE: March 24, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/195
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 120
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-153-051B-28

Query Match 0.8%; Score 36; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 70 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 105

RESULT 27
US-08-060-952C-44
Sequence 44, Application US/08060952C
Patent No. 5695932
GENERAL INFORMATION:
APPLICANT: Michael D. West
APPLICANT: Jerry W. Shay
APPLICANT: Woodring E. Wright
APPLICANT: Elizabeth Blackburn
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF CONDITIONS
TITLE OF INVENTION: RELATED TO TELOMERE LENGTH AND/OR
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/060,952C
FILING DATE: May 13, 1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/882,438

FILING DATE: May 13, 1992
APPLICATION NUMBER: 08/038,766
FILING DATE: March 24, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 202/045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 120
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-060-952C-44

Query Match 0.8%; Score 36; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 70 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 105

RESULT 28
US-08-151-477A-28
Sequence 28, Application US/08151477A
Patent No. 5830644
GENERAL INFORMATION:
APPLICANT: Michael D. West
APPLICANT: Jerry W. Shay
APPLICANT: Woodring E. Wright
APPLICANT: Elizabeth Blackburn
APPLICANT: Nam Woo Kim
APPLICANT: Calvin B. Harley
APPLICANT: Scott L. Weinlich
APPLICANT: Catherine Strahl
APPLICANT: Michael J. McEachern
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF
TITLE OF INVENTION: CONDITIONS RELATED TO TELOMERE
TITLE OF INVENTION: LENGTH AND/OR TELOMERE ACTIVITY
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/151,477A
FILING DATE: No. 5830644ember 12, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/038,766
FILING DATE: March 24, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 202/189
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 120
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-151-477A-28

Query Match 0.8%; Score 36; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4337 TGTAAATAAAAAAAAAAAAAAAAAAAAAA 4372
Db 70 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 105

RESULT 29
US-08-819-867-58
Sequence 58, Application US/08819867
Patent No. 6007989
GENERAL INFORMATION:
APPLICANT: Michael D. West
APPLICANT: Calvin B. Harley
APPLICANT: Scott L. Weinrich
APPLICANT: Catherine M. Strahl
APPLICANT: Michael J. McEachern
APPLICANT: Jerry Shay
APPLICANT: Woodring E. Wright
APPLICANT: Elizabeth H. Blackburn
APPLICANT: Nam Woo Kim
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF
TITLE OF INVENTION: CONDITIONS RELATED TO
TITLE OF INVENTION: TELOMERE LENGTH AND/OR
TITLE OF INVENTION: TELOMERASE ACTIVITY
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/819,867
FILING DATE: March 14, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/153,051
FILING DATE: No. 6007989 September 12, 1993
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Chambers, Daniel M.
REGISTRATION NUMBER: 34,561
REFERENCE/DOCKET NUMBER: 224/232
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 120 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-819-867-58

Query Match 0.8%; Score 36; DB 3; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4337 TGTAAATAAAAAAAAAAAAAAAAAAAAAA 4372
Db 70 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 105

RESULT 30
US-08-464-011B-44
Sequence 44, Application US/08464011B
Patent No. 6368789
GENERAL INFORMATION:
APPLICANT: Michael D. West
APPLICANT: Jerry W. Shay
APPLICANT: Woodring E. Wright
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF CONDITIONS
RELATED TO TELOMERE LENGTH AND/OR
TELOMERASE ACTIVITY
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,011B
FILING DATE: 05-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/882,438
FILING DATE: May 13, 1992
APPLICATION NUMBER: 08/038,766
FILING DATE: March 24, 1993
APPLICATION NUMBER: 08/060,952
FILING DATE: May 13, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Walburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 202/045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 120
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-08-464-011B-44
Query Match 0.8%; Score 36; DB 3; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db      70 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 105
|||||
RESULT 31
US-09-378-535-58
; Sequence 58, Application US/09378535
; Patent No. 6551774
; GENERAL INFORMATION:
; APPLICANT: Michael D. West
; Calvin B. Harley
; Scott L. Weinrich
; Catherine M. Strahl
; Michael J. McEachern
; Jerry Shay
; Elizabeth H. Blackburn
; Nam Woo Kim
; Homaoun Vaziri
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF
; CONDITIONS RELATED TO
; TEOLOMERE LENGTH AND/OR
; TELOMERASE ACTIVITY
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Filth Street
; Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S. A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/378,535
; FILING DATE: 20-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/819,867
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Chambers, Daniel M.
; REGISTRATION NUMBER: 34,561
; REFERENCE/DOCKET NUMBER: 224/232
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-09-378-535-58

Query Match      0.8%; Score 36; DB 4; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
|||||
Db      70 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 105
|||||

RESULT 32
US-09-621-976-15316
; Sequence 15316, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jober, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15316
; LENGTH: 168
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-15316

Query Match      0.8%; Score 36; DB 4; Length 168;
Best Local Similarity 100.0%; Pred. No. 4.1e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
|||||
Db      132 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 167
|||||

RESULT 33
US-09-621-976-15319
; Sequence 15319, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jober, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15319
; LENGTH: 173
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-15319

Query Match      0.8%; Score 36; DB 4; Length 173;
Best Local Similarity 100.0%; Pred. No. 4.1e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
|||||
Db      132 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 167
|||||

RESULT 34
US-09-621-976-15322
; Sequence 15322, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jober, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15322
; LENGTH: 178
; TYPE: DNA
```

```

; ORGANISM: Homo sapiens
US-09-621-976-15322

Query Match
Best Local Similarity 100.0%; Score 36; DB 4; Length 178;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 43372
DB 132 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 167

RESULT 35
US-09-621-976-15320
; Sequence 15320, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15320
; LENGTH: 180
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-15320

Query Match
Best Local Similarity 100.0%; Score 36; DB 4; Length 180;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 43372
DB 132 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 167

RESULT 36
US-09-621-976-15317
; Sequence 15317, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15317
; LENGTH: 194
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-15317

Query Match
Best Local Similarity 100.0%; Score 36; DB 4; Length 194;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 43372
DB 132 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 167

RESULT 37
US-09-792-594-11
; Sequence 11, Application US/09792594

```

```

; Patent NO. 6436706
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Walt
; TITLE OF INVENTION: ANTISENSE MODULATION OF RECOL4 EXPRESSION
; FILE REFERENCE: RTS-0209
; CURRENT APPLICATION NUMBER: US/09/792,594
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 11
; LENGTH: 223
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-792-594-11

Query Match
Best Local Similarity    100.0%; Score 36; DB 3; Length 223;
Matches   36; Conservative    0; Mismatches      0; Indels      0; Gaps      0;

Cy       4337 TGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
          |||||||
Db        167 TGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 202

RESULT 38
US-09-621-976-2841
; Sequence 2841, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jodert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2841
; LENGTH: 284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 21..182
US-09-621-976-2841

Query Match
Best Local Similarity    100.0%; Score 36; DB 4; Length 284;
Matches   36; Conservative    0; Mismatches      0; Indels      0; Gaps      0;

Cy       4337 TGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
          |||||||
Db        245 TGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 280

RESULT 39
US-09-949-016-185724/c
; Sequence 185724, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498

```

```

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 185724
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-185724

Query Match
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4371
DB 83 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 48

RESULT 40
US-09-800-729-21
; Sequence 21, Application US/09800729
; Patent No. 6605592
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-800-729-21

Query Match
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 630 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 665

RESULT 41
US-09-621-976-2461
; Sequence 2461, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2461
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 64..552
; NAME/KEY: misc_feature
; LOCATION: 33
; OTHER INFORMATION: n=a, g, c or t
,
```

```

US-09-621-976-2461

Query Match
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
DB 591 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 626

RESULT 42
US-09-907-794A-376
; Sequence 376, Application US/09907794A
; Patent No. 6635468
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,794A
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
,
```



```

; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 376
; LENGTH: 997
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-907-794A-376

Query Match      0.8%; Score 36; DB 4; Length 997;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      915 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 950

RESULT 43
US-09-905-125A-376
; Sequence 376, Application US/09905125A
; Patent No. 6664376
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mathew, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,125A
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
```

```

; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 376
; LENGTH: 997
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-905-125A-376

Query Match      0.8%; Score 36; DB 4; Length 997;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      915 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 950

RESULT 44
US-09-902-775A-376
; Sequence 376, Application US/09902775A
; Patent No. 666451
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mathew, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,775A
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
```

PRIOR FILING DATE: 1999-07-07
 PRIOR APPLICATION NUMBER: US 60/145,698
 PRIOR FILING DATE: 1999-07-26
 PRIOR APPLICATION NUMBER: US 60/146,222
 PRIOR FILING DATE: 1999-07-28
 PRIOR APPLICATION NUMBER: PCT/US99/20594
 PRIOR FILING DATE: 1999-09-08
 PRIOR APPLICATION NUMBER: PCT/US99/20944
 PRIOR FILING DATE: 1999-09-13
 PRIOR APPLICATION NUMBER: PCT/US99/21090
 PRIOR FILING DATE: 1999-09-15
 PRIOR APPLICATION NUMBER: PCT/US99/21547
 PRIOR FILING DATE: 1999-09-15
 PRIOR APPLICATION NUMBER: PCT/US99/23089
 PRIOR FILING DATE: 1999-10-05
 PRIOR APPLICATION NUMBER: PCT/US99/28214
 PRIOR FILING DATE: 1999-11-29
 PRIOR APPLICATION NUMBER: PCT/US99/28313
 PRIOR FILING DATE: 1999-11-30
 PRIOR APPLICATION NUMBER: PCT/US99/28564
 PRIOR FILING DATE: 1999-12-02
 PRIOR APPLICATION NUMBER: PCT/US99/28565
 PRIOR FILING DATE: 1999-12-02
 PRIOR APPLICATION NUMBER: PCT/US99/30095
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: PCT/US99/30911
 PRIOR FILING DATE: 1999-12-20
 PRIOR APPLICATION NUMBER: PCT/US99/30999
 PRIOR FILING DATE: 1999-12-20
 PRIOR APPLICATION NUMBER: PCT/US00/00219
 PRIOR FILING DATE: 2000-01-05
 NUMBER OF SEQ ID NOS: 423
 SEQ ID NO 376
 LENGTH: 997
 TYPE: DNA
 ORGANISM: Homo Sapien
 US-09-902-775A-376

Query Match 0.8%; Score 36; DB 4; Length 997;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4337 TGTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
 DB 915 TGTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 950

RESULT 45
 US-09-906-700-376
 Sequence 376, Application US/09906700
 Patent No. 6723535
 GENERAL INFORMATION:
 APPLICANT: Genentech, Inc.
 APPLICANT: Ashkenazi, Avi
 APPLICANT: Botstein, David
 APPLICANT: Desnoyers, Luc
 APPLICANT: Eaton, Dan L.
 APPLICANT: Ferrara, Napoleone
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Fong, Sherman
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gerber, Hanspeter
 APPLICANT: Gerlitsen, Mary E.
 APPLICANT: Goddard, A.
 APPLICANT: Grimaldi, Christopher J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Hillen, Kenneth, J.
 APPLICANT: Kijavlin, Ivar J.
 APPLICANT: Mather, Jennie P.
 APPLICANT: Pan, James
 APPLICANT: Paoni, Nicholas F.
 APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.
 APPLICANT: Tumas, Daniel
 APPLICANT: Williams, P. Mickey
 APPLICANT: Wood, William, I.
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 FILE REFERENCE: 10466-14
 CURRENT APPLICATION NUMBER: US/09/906,700
 CURRENT FILING DATE: 2000-09-18
 PRIOR APPLICATION NUMBER: PCT/US00/04414
 PRIOR FILING DATE: 2000-02-22
 PRIOR APPLICATION NUMBER: US 60/143,048
 PRIOR FILING DATE: 1999-07-07
 PRIOR APPLICATION NUMBER: US 60/145,698
 PRIOR FILING DATE: 1999-07-26
 PRIOR APPLICATION NUMBER: US 60/146,222
 PRIOR FILING DATE: 1999-07-28
 PRIOR APPLICATION NUMBER: PCT/US99/20594
 PRIOR FILING DATE: 1999-09-08
 PRIOR APPLICATION NUMBER: PCT/US99/20944
 PRIOR FILING DATE: 1999-09-13
 PRIOR APPLICATION NUMBER: PCT/US99/21090
 PRIOR FILING DATE: 1999-09-15
 PRIOR APPLICATION NUMBER: PCT/US99/21547
 PRIOR FILING DATE: 1999-09-15
 PRIOR APPLICATION NUMBER: PCT/US99/23089
 PRIOR FILING DATE: 1999-10-05
 PRIOR APPLICATION NUMBER: PCT/US99/28214
 PRIOR FILING DATE: 1999-11-29
 PRIOR APPLICATION NUMBER: PCT/US99/28313
 PRIOR FILING DATE: 1999-11-30
 PRIOR APPLICATION NUMBER: PCT/US99/28564
 PRIOR FILING DATE: 1999-12-02
 PRIOR APPLICATION NUMBER: PCT/US99/28565
 PRIOR FILING DATE: 1999-12-02
 PRIOR APPLICATION NUMBER: PCT/US99/30095
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: PCT/US99/30911
 PRIOR FILING DATE: 1999-12-20
 PRIOR APPLICATION NUMBER: PCT/US99/30999
 PRIOR FILING DATE: 1999-12-20
 PRIOR APPLICATION NUMBER: PCT/US00/00219
 PRIOR FILING DATE: 2000-01-05
 NUMBER OF SEQ ID NOS: 423
 SEQ ID NO 376
 LENGTH: 997
 TYPE: DNA
 ORGANISM: Homo Sapien
 US-09-906-700-376

Query Match 0.8%; Score 36; DB 4; Length 997;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4337 TGTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
 DB 915 TGTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 950

RESULT 46
 US-09-903-603A-376
 Sequence 376, Application US/09903603A
 Patent No. 6767995
 GENERAL INFORMATION:
 APPLICANT: Genentech, Inc.
 APPLICANT: Ashkenazi, Avi
 APPLICANT: Botstein, David
 APPLICANT: Desnoyers, Luc
 APPLICANT: Eaton, Dan L.
 APPLICANT: Ferrara, Napoleone
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Fong, Sherman
 APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: GNE.1618P2C12
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: US/09/903, 603A
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/143, 048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145, 698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146, 222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28213
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 376
LENGTH: 997
TYPE: DNA
ORGANISM: Homo Sapien
US-09-903-603A-376

Query Match 0.8%; Score 36; DB 4; Length 997;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 915 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 950

RESULT 47
US-09-904-920A-376

Sequence 376, Application US/09904920A
Patent No. 6806352
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143, 048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145, 698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146, 222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28213
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 376
LENGTH: 997
TYPE: DNA
ORGANISM: Homo Sapien
US-09-904-920A-376

Query Match 0.8%; Score 36; DB 4; Length 997;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 915 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 950

RESULT 48
US-09-909-064-376
; Sequence 376, Application US/09909064
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Baton, Dan L.
; APPLICANT: Ferreira, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Thomas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,064
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 376
; LENGTH: 997
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-909-064-376

Query Match 0.8%; Score 36; DB 4; Length 997;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 915 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 950

RESULT 49
US-09-905-381A-376
; Sequence 376, Application US/09905381A
; Patent No. 6818746
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Baton, Dan L.
; APPLICANT: Ferreira, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Thomas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,381A
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15

```

; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 376
; LENGTH: 997
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-09-905-381A-376

Query Match          0.8%; Score 36; DB 4; Length 997;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      915 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 950

RESULT 50
US-09-906-618-376
; Sequence 376, Application US/09906618
; Patent No. 6828146
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerltsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie F.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,618
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
```

```

; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 376
; LENGTH: 997
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-09-906-618-376

Query Match          0.8%; Score 36; DB 4; Length 997;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db      915 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 950

RESULT 51
US-09-522-714-1
; Sequence 1, Application US/09522714
; Patent No. 6563020
; GENERAL INFORMATION:
; APPLICANT: Stimmens, Carl R.
; APPLICANT: Yaipani, Nasser
; TITLE OF INVENTION: Maize Chitinases and Their Use in
; FILE REFERENCE: 1100
; CURRENT APPLICATION NUMBER: US/09/522,714
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 60/125,915
; EARLIER FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (51)...(893)
; US-09-522-714-1

Query Match          0.8%; Score 36; DB 4; Length 1094;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
```

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1050 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1085

RESULT 52

US-09-152-060-15
; Sequence 15, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P1.US
; CURRENT APPLICATION NUMBER: US/09/152,060
; EARLIER FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,357
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,189
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/057,765
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/068,368
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1123
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-152-060-15

Query Match 0.8%; Score 36; DB 3; Length 1123;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1049 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1084

RESULT 53

US-09-436-521A-5
; Sequence 5, Application US/09436521A
; Patent No. 6583337
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Glucose-6-Phosphate/Phosphate Translocator
; FILE REFERENCE: B81269 US NA
; CURRENT APPLICATION NUMBER: US/09/436,521A
; EARLIER FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: 60/107,910
; PRIOR FILING DATE: 1998-11-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 1281
; TYPE: DNA
; ORGANISM: Triticum aestivum

US-09-436-521A-5

Query Match 0.8%; Score 36; DB 4; Length 1281;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1229 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1264

RESULT 54

US-09-499-302A-1
; Sequence 1, Application US/09499302A
; Patent No. 6369212
; GENERAL INFORMATION:
; APPLICANT: BOUNG-JUN, OH
; APPLICANT: MOON, KYUNG KO
; APPLICANT: YOUNG, SOON KIM
; TITLE OF INVENTION: A CYTOCHROME P450 GENE HIGHLY EXPRESSED IN THE
; FILE REFERENCE: 10324/P6443USO
; CURRENT APPLICATION NUMBER: US/09/499,302A
; EARLIER FILING DATE: 2000-02-07
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1781
; TYPE: DNA
; ORGANISM: Capsicum annuum
US-09-499-302A-1

Query Match 0.8%; Score 36; DB 3; Length 1781;
Best Local Similarity 100.0%; Pred. No. 3.8e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1727 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1762

RESULT 55

US-08-486-049-1
; Sequence 1, Application US/08486049
; Patent No. 6572862
; GENERAL INFORMATION:
; APPLICANT: Estes, Mary K
; APPLICANT: Jiang, Xi
; APPLICANT: Graham, David Y
; TITLE OF INVENTION: Methods and Reagents to Detect and
; TITLE OF INVENTION: Characterize No. 6572862walk and Related Viruses
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 801 Pennsylvania Ave., N.W.
; CITY: Washington, D.C.
; STATE:
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,049
; FILING DATE: June 7, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Davis, Peter
; REGISTRATION NUMBER: 36,119
; REFERENCE/DOCKET NUMBER: 311.023
; TELECOMMUNICATION INFORMATION:

```
?
? TELEPHONE: 202-662-0200
? TELEFAX: 202-662-4643
?
? TELEX:
?
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 7724 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: unknown
? MOLECULE TYPE: CDNA
? ORIGINAL SOURCE:
? ORGANISM: No. 6572862walk virus
? STRAIN: 8P1ta
? INDIVIDUAL ISOLATE: 8P1ta
? IMMEDIATE SOURCE:
? CLONE: pUCNV-953 and its derivatives
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 146..5359
? OTHER INFORMATION: /note= "The protein encoded by
? OTHER INFORMATION: nucleotides 146 through 5359 is eventually cleaved
? OTHER INFORMATION: to make at least a picornavirus 2c-like protein, a
? OTHER INFORMATION: 3C-like protease and an RNA-dependent RNA polymerase."
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 5346..6935
? OTHER INFORMATION: /note= "Nucleotides 5346 through
? OTHER INFORMATION: 5359 are used for coding two different amino acid
? OTHER INFORMATION: sequences: the first is the sequence coded by
? OTHER INFORMATION: nucleotides 146 through 5359, the second by nucleotides
? OTHER INFORMATION: 5346 through 6935."
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 6938..7573
? US-08-486-049-1
```

```

Query Match Similarity      0.8%; Score 36; DB 4; Length 7724;
Best Local Similarity      100.0%; Pred. No. 3.6e-05;
Matches    36; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

QY          4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
            |||||
Db           7639 TGTTAAAAAAAAAAAAAAAAAAAAAAAAAA 7674

RESULT 56
US-09-949-016-17054/c
; Sequence 17054, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17054
; LENGTH: 28696
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(28696)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17054
```

Query Match 0.8%; Score 36; DB 4; Length 28696;
 Best Local Similarity 100.0%; Pred. No. 3.5e-05;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

```

RESULT 57
US-09-949-016-12335/c
: Sequence 12335, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: TITLE OR INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CL001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 12335
: LENGTH: 28780
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(28780)
: OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12335

```

```

Query Match      0.8%; Score 36; DB 4; Length 28780;
                0.8%; Score 36; DB 4; Length 28780;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          4336 CTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4371
              |||||
Db           5056 CTGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5021

RESULT 58
US-09-621-976-14003
; Sequence 14003, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 14003
; LENGTH: 115
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12
; OTHER INFORMATION: n=a, g, c or t

US-09-621-976-14003

Query Match      0.8%; Score 35; DB 4; Length 115;
```

Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAA
Db 66 GTTAAAAA
100

RESULT 59
US-09-621-976-13418

; Sequence 13418, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 13418
; LENGTH: 187
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-13418

Query Match 0.8%; Score 35; DB 4; Length 187;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAA
Db 145 GTTAAAAA
179

RESULT 60
US-08-248-016-11

; Sequence 11, Application US/08248016
; Patent No. 5550109
; GENERAL INFORMATION:
; APPLICANT: Schonwetter, Barry S.
; APPLICANT: Zaslloff, Michael A.
; TITLE OF INVENTION: Inducible Defensin Peptide From
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESS: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/248.016
; FILING DATE: 24-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ogden, Stasia L.
; REGISTRATION NUMBER: 36,228
; REFERENCE/DOCKET NUMBER: 05387.0017-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-248-016-11

Query Match 0.8%; Score 35; DB 1; Length 350;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAA
Db 315 GTTAAAAA
349

RESULT 61
US-08-451-501-11

; Sequence 11, Application US/08451501
; Patent No. 5656738
; GENERAL INFORMATION:
; APPLICANT: Schonwetter, Barry S.
; APPLICANT: Zaslloff, Michael A.
; TITLE OF INVENTION: Inducible Defensin Peptide From
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESS: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451.501
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/248.016
; FILING DATE: 24-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 30,907
; REFERENCE/DOCKET NUMBER: 05387.0017-01000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-451-501-11

Query Match 0.8%; Score 35; DB 1; Length 350;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAA
Db 315 GTTAAAAA
349

RESULT 62

PCT-US95-06761-11
; Sequence 11, Application PC/TUS9506761
; GENERAL INFORMATION:
; APPLICANT: Magalini Pharmaceuticals Inc.
; APPLICANT: 5110 Campus Drive
; APPLICANT: Plymouth Meeting, PA 19462
; TITLE OF INVENTION: Inducible Defense Peptide From
; TITLE OF INVENTION: Mammalian Epithelia
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06761
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/248,016
; FILING DATE: 24-MAY-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ogden, Stasia L.
; REGISTRATION NUMBER: 36,228
; REFERENCE/DOCKET NUMBER: 05387.0017-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US95-06761-11

Query Match 0.8%; Score 35; DB 5; Length 350;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 315 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 349

RESULT 63
US-09-621-976-16710
; Sequence 16710, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 16710
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-16710

Query Match 0.8%; Score 35; DB 4; Length 413;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 375 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 409

RESULT 64
US-09-621-976-16712
; Sequence 16712, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 16712
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-16712

Query Match 0.8%; Score 35; DB 4; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 377 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 411

RESULT 65
US-09-621-976-16711
; Sequence 16711, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 16711
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-16711

Query Match 0.8%; Score 35; DB 4; Length 430;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 375 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 409

RESULT 66
US-09-621-976-2632
; Sequence 2632, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:

```

; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2632
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 147..335
; OTHER INFORMATION: n = A,T,C or G
; US-09-621-976-2632

Query Match
Best Local Similarity 100.0%; Score 35; DB 4; Length 469;
Pred. No. 0.0001; Mismatches 0; Indels 0; Gaps 0;
Matches 35; Conservative 0;

Qy 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 431 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 465

RESULT 67
US-09-621-976-2630
; Sequence 2630, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2630
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 147..335
; NAME/KEY: misc_feature
; LOCATION: 474
; OTHER INFORMATION: n=a, g, c or t
; US-09-621-976-2630

Query Match
Best Local Similarity 100.0%; Score 35; DB 4; Length 475;
Pred. No. 0.0001; Mismatches 0; Indels 0; Gaps 0;
Matches 35; Conservative 0;

Qy 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 431 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 465

RESULT 68
US-09-385-982-537
; Sequence 537, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGS, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
```

```

; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: PatSeq for windows Version 3.0
; SEQ ID NO 537
; LENGTH: 794
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(794)
; OTHER INFORMATION: n = A,T,C or G
; US-09-385-982-537
```

```

Query Match
Best Local Similarity 100.0%; Score 35; DB 3; Length 794;
Pred. No. 0.0001; Mismatches 0; Indels 0; Gaps 0;
Matches 35; Conservative 0;
```

```

Qy 4337 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4371
Db 112 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 146
```

```

RESULT 69
US-10-039-659A-5
; Sequence 5, Application US/10039659A
; Patent No. 6723520
; GENERAL INFORMATION:
; APPLICANT: Wang, Wei
; APPLICANT: Gish, Kurt C.
; APPLICANT: Schall, Thomas J.
; APPLICANT: Vicari, Alain P.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Antibodies that bind chemokine TECK
; FILE REFERENCE: DX0589XIB US
; CURRENT APPLICATION NUMBER: US/10/039,659A
; CURRENT FILING DATE: 2002-01-03
; PRIOR APPLICATION NUMBER: US 08/887,977
; PRIOR FILING DATE: 1997-07-03
; PRIOR APPLICATION NUMBER: US 60/021,664
; PRIOR FILING DATE: 1996-07-05
; PRIOR APPLICATION NUMBER: US 60/028,329
; PRIOR FILING DATE: 1996-10-11
; PRIOR APPLICATION NUMBER: US 60/048,593
; PRIOR FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(288)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (79)..()
; OTHER INFORMATION:
; US-10-039-659A-5
```

```

Query Match
Best Local Similarity 100.0%; Score 35; DB 4; Length 801;
Pred. No. 0.0001; Mismatches 0; Indels 0; Gaps 0;
Matches 35; Conservative 0;
```

```

Qy 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 762 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 796
```

```
RESULT 70
US-09-318-448-18
; Sequence 18, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenicos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-18

Query Match
0.8%; Score 35; DB 3; Length 1192;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1151 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1185

RESULT 71
US-09-614-221A-3
; Sequence 3, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunananda, Balasubramanian
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; TITLE OF INVENTION: WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 3
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (32)...(1099)
US-09-614-221A-3

Query Match
0.8%; Score 35; DB 4; Length 1355;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1299 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1333

RESULT 72
US-09-501-115-5
; Sequence 5, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Pader, Gary M.
; APPLICANT: Rafalecki, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
```

```
; FILE REFERENCE: BBI328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; CURRENT FILING DATE: 2000-02-09
; EARLIER APPLICATION NUMBER: 60/119,585
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 1414
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-501-115-5

Query Match
0.8%; Score 35; DB 4; Length 1414;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1319 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1353

RESULT 73
US-09-039-046-1
; Sequence 1, Application US/09039046
; Patent No. 6331660
; GENERAL INFORMATION:
; APPLICANT: CHOMET, PAUL S.
; APPLICANT: FREY, MONIKA
; APPLICANT: GIERL, ALFONS
; TITLE OF INVENTION: MAIZE DIMBOA BIOSYNTHESIS GENES
; FILE REFERENCE: DKEM:131
; CURRENT APPLICATION NUMBER: US/09/039,046
; CURRENT FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1508
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (49)..(1089)
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Primer
US-09-039-046-1

Query Match
0.8%; Score 35; DB 3; Length 1508;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
Db 1411 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1445

RESULT 74
US-09-205-258-40
; Sequence 40, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
```

; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,880
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,896
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/049,020
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,876
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,895
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,884
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,894
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,971
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,964
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,882
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,899
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,893
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,900
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,901
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,892
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,915
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/049,019
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,970
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,972
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,916
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/049,373
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,875
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/049,374
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,917
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,949
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,974
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,883
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,897
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,898
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,962
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,963
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,877
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/048,878
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/070,923
 ; EARLIER FILING DATE: 1997-12-18
 ; EARLIER APPLICATION NUMBER: 60/092,921
 ; EARLIER FILING DATE: 1998-07-15
 ; EARLIER APPLICATION NUMBER: 60/094,657
 ; EARLIER FILING DATE: 1998-07-30

; NUMBER OF SEQ ID NOS: 1227
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO: 40
 ; LENGTH: 1515
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (69)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; US-09-205-258-40

Query Match 0.8%; Score 35; DB 4; Length 1515;
 Best Local Similarity 100.0%; Pred. No. 0.0001;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 ;
 Qy 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
 Db 1476 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1510

RESULT 75
 US-08-665-966-9
 ; Sequence 9, Application US/08665966
 ; Patent No. 5756328
 ; GENERAL INFORMATION:
 ; APPLICANT: Steffens, John C.
 ; APPLICANT: Ghangas, Gurdev S.
 ; TITLE OF INVENTION: Acyl Transferase and Gene Encoding Acyl
 ; TITLE OF INVENTION: Transferase
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Jones, Tullar & Cooper, P.C.
 ; STREET: P.O. Box 2266 Eads Station
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: USA
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentln Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/665,966
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Spector, Eric S.
 ; REGISTRATION NUMBER: 22495
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-415-1500
 ; TELEFAX: 703-415-1508
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1604 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 55..1392
 ; US-08-665-966-9

Query Match 0.8%; Score 35; DB 1; Length 1604;
 Best Local Similarity 100.0%; Pred. No. 0.0001;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 ;
 Qy 4338 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4372
 Db 1546 GTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1580

Search completed: March 25, 2005, 14:34:04
Job time : 686 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2005, 15:22:46 ; Search time 43 Seconds
(without alignments)
1373.886 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3238
Sequence: 1 MTRSAVFAKAAAPAGANPE.....RSRCRLIOKKQTPHRTCRK 614

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	211	6.5	955	2 S52959	male-specific leth
2	202	6.2	854	2 S02003	neurofilament trip
3	200	6.2	3938	2 T42761	Bassoon protein
4	199	6.1	1020	1 QFH0H	neurofilament trip
5	197.5	6.1	1072	1 A37221	neurofilament trip
6	197.5	6.1	3942	2 T42730	Bassoon protein
7	192	5.9	3534	2 T42567	legumen protein 2
8	190	5.9	1087	1 QPM5H	neurofilament trip
9	190	5.9	1098	2 A43459	probable transcript
10	189.5	5.9	606	2 T28770	neurofilament trip
11	189	5.8	539	2 T30427	hypothetical prote
12	188	5.8	2715	2 T13049	eyelid - fruit fly
13	186.5	5.8	460	2 T33110	hypothetical prote
14	184.5	5.7	2649	2 T51023	hypothetical prote
15	183.5	5.7	1520	2 T00273	hypothetical prote
16	182	5.6	1216	2 T34101	hypothetical prote
17	181.5	5.6	2187	2 T30826	nascent polypeptid
18	178	5.5	1585	2 T31611	hypothetical prote
19	176.5	5.5	7962	2 T38346	elastic titin - hu
20	173	5.3	351	1 OZ20K	circumsporozoite p
21	173	5.3	1207	2 T00378	glucan 41 protein -
22	173	5.3	2022	2 T48818	PTB-associated spl
23	172.5	5.3	707	2 A46302	extensin-like prot
24	172.5	5.3	1188	2 S49915	high molecular mas
25	172	5.3	1151	2 T18535	MHC class III hist
26	172	5.3	1870	2 S37671	MHC class III hist
27	172	5.3	1872	2 S36152	MHC class III hist
28	171.5	5.3	3968	2 A44265	trithorax homolog
29	171	5.3	1285	2 T14171	ataxin-2 - mouse

30	171	5.3	1647	2 S45252	SNF2beta protein -
31	170.5	5.3	418	2 T15142	hypothetical prote
32	170	5.3	834	2 T42702	hypothetical prote
33	169.5	5.2	858	2 S15762	neurofilament trip
34	169.5	5.2	891	2 G84693	probable prolina-r
35	169	5.2	268	2 A49303	homeotic protein C
36	169	5.2	1613	2 S39059	protein BRG1 - hum
37	169	5.2	2342	2 T13412	hypothetical prote
38	168.5	5.2	3530	2 A59266	unconventional myo
39	167	5.2	398	2 T34947	hypothetical prote
40	167	5.2	657	2 B84869	probable Spf6 prot
41	167	5.2	1320	2 UC5630	TCOF1 protein - mo
42	167	5.2	2142	2 B35098	MHC class III hist
43	166.5	5.1	6642	2 T29757	protein UNC-89 - C
44	166	5.1	753	2 J00532	OP protein - Kenne
45	166	5.1	1634	2 T26517	hypothetical prote

ALIGNMENTS

RESULT 1

S52959 male-specific lethal-1 protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004

C:Accession: S52959; S65350

R:Palmer, M.U.; Meigner, V.A.; Richman, R.; Manning, J.E.; Kuroda, M.I.; Lucchesi, J.C.

Genetics 134, 545-557, 1993

A:Title: The male-specific lethal-one (msl-1) gene of Drosophila melanogaster encodes a

A:Reference number: S52959; WUID:93314941; PMID:8325488

A:Accession: S52959

A:Molecule type: DNA

A:Residues: 1955 <PAL>

A:Cross-references: UNIPROT:P50535; EMBL:L14582

R:Palmer, M.U.; Meigner, V.A.; Richman, R.; Manning, J.E.; Kuroda, M.I.; Lucchesi, J.C.

submitted to the EMBL Data Library, April 1993

A:Reference number: S65350

A:Accession: S65350

A:Molecule type: DNA

A:Residues: 1407, 'S', 408-585, 'M', 587-955 <PAM>

A:Cross-references: EMBL:L14582

C:Genetics:

A:Gene: msl-1

A:Cross-references: FlyBase:FBgn0005617

A:Insertions: 329/3

Query Match 6.5%; Score 211; DB 2; Length 955;

Best Local Similarity 20.3%; Pred. No. 0.0072;

Matches 96; Conservative 91; Mismatches 171; Indels 114; Gaps 18;

QY	181	PPLAPTRAGTTLAASGRKMSRKSPGGGSGAGSQAACLKQIL	-----LQDLIE 234
DB	485	PKHLPRVAVPKVTKTSRSTLPK-----NTADIKDPAKVAHNMSTQTDPVK	537
QY	235	QOOOQLOAKKEI-----ELKSERDT-LIARIRMRBMQLVKNCKEKHKL	282
DB	538	TORLOVKIKROYENHPDMRTGSSAPSDIRKOKANDPVPSTPFTKTKSKSLVNDKTTSET	597
QY	283	FOGYETEERRET---ELSEKIKLECOPELSETSQTLP-----PKPSCGRSGK	327
DB	598	SQSPDQIDIVETRYRKLALHKLKELLSQSHSQVTLKIKREAVATNLIYPP-----	649
QY	328	GHRKSPGSGTEKTPYKLAPEFSKYTKTPG---SPIKEPCCGSLSTVCKRELRSQ	384
DB	650	---SAPVST---TTTPAPTPSTTPPGSTPHAVTSSMDQISAKSKSKAAEQI---	699
QY	385	ETPEKPPSSVDTPEPLSTPOK-----GPSTHPPEKAFSSRIED-----LPYLSTT	429
DB	700	ATPLTPQSNSSVSSTSTIRKTLNMGSPHTYSATATARSGLQGRFRATPTPTSTRTWEDQ	759
QY	430	-----EMYLGRWHPPPSPPLPLRESSPKKETVARCLMPSVAGETSVLAVPSWRDHSV	483

QY 534 LQLRMVKKKGIOSESEVTSFPEPDDVSLMTPPLPVVAFGRPLPKLTPONFELPMWD 593
 DB 860 VSGRGLAKGGAQK-----ASARPPPESSQSEVALP--- 889
 QY 594 ERSRCRLIEIQKQTPH 609
 DB 890 -----KRLPH 895

RESULT 4

QFHU
 neurofilament triplet H protein - human
 N:Alternate names: neurofilament protein, 112K
 C:Species: Homo sapiens (man)
 C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
 C/Accession: S00979
 R:Leees, J.F.; Sheldman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.
 EMO J. 7, 1947-1955, 1988
 A>Title: The structure and organization of the human heavy neurofilament subunit (NF-H)
 A/Reference number: S00979; PMID:8838981; PMID:3138108
 A/Accession: S00979
 A:Molecule type: DNA
 A/Residues: 1-1020 <LEB>
 A/Cross-references: UNIPROT:P12036; EMBL:X15306; NID:G35028; PIDN:CAA33366.1; PID:G18414
 A/Note: It is uncertain whether Met-1 or Met-2 is the initiator
 C/Genetics:
 A:Gene: GDB:NEFH
 A/Cross-references: GDB:120225; OMIM:162230
 A:Map position: 22q12.1-22q13.1
 A:Intons: 295/1; 361/3; 403/2
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
 F:1-100/Domain: amino-terminal <NTD>
 F:101-410/Domain: rod #status predicted <ROD>
 F:411-1020/Domain: carboxyl-terminal <CTD>
 F:502-826/Region: 14-residue repeats
 F:503, 511, 518, 526, 532, 540, 546, 552, 560, 566, 574, 580, 586, 594, 600, 606, 614, 620, 628, 634, 640, 644
 (covalent) #status predicted
 F:732, 768/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 6.1%; Score 199; DB 1; Length 1020;
 Best Local Similarity 23.6%; Pred. No. 0.028;
 Matches 119; Conservative 70; Mismatches 230; Indels 86; Gaps 22;

QY 90 GQGESEWGSVPLPCPPATKQIGEPAAAGSPRRKQAVLP-OTGLVAAKE 148
 DB 489 GEEHEAEGGEETKSP-----AEEAAS-----PEKAESPVEKAESPAAKS 532
 QY 149 PTWAGDKGAASPAAT-----ASDA-----GPPPLPLPPLPPLAATATAGTLAASGR 198
 DB 533 P-----EKEAKSPAAVKSPEKAKSPKAEKASPPKAEKASPAVKSPEKAKSP 567
 QY 199 WKSMRKSPLGGGSGGASQAACIKQILLQLDLIEQQOOLQAKETELKSERDTLL 258
 DB 588 AKKEAASPAEAKSPKAEKSPV-----BEAKSPAAKSPVVEAKSP----- 629
 QY 259 ARIERHERMQLVKQDNEKRRHLFOGYETEEBEETELSKITLCOPELSTISQTLPPK 318
 DB 630 AEVKSPEK-----AKSPTEEAKSPKAEKSPKAEKASPAVKAESPAAKS 684
 QY 319 PFSGSGSGKGRKSPGSEKTERKTPVKLAPEFSKYTKTPKHSPIKEE-PCGSLSTVC 377
 DB 685 K---AAAKSEKAKSPV-KEAKSPKAEKASPAVKAESPAAKS 740
 QY 378 KRELRSQE--TEKPRSSVDTPEPRLSTPOKSPSTHP---KKAFFSEIEDLPYLSTTE 430
 DB 741 KEBAKSPKAEKSPKAEKATLDVKSPEAKTPAKBEARSPADKFPKAEKSPVEEVKSPKAK 800
 QY 431 MYLCRMHQPSPPLPLRE--SSPKER--TVARCLMPSSVAGTSTLAVPSM---RDHS 482
 DB 801 SPLKADAKAEKEIKPKKEEVKSPVKEBEKQEVKVEPKKAEKAPATPKTEKKDSK 860

QY 483 VE--PLRDNPQDLLENLDSVFSKHALELDEKRRKMDIORT--RECRILQILRM 538
 DB 861 KEAPKPEKAPKPKVEKKEKPAVEKPKESKVE--AKKEAEADKKVTPPEKAPAKVEV-- 916
 QY 539 YKKKGIOSESEPVTSFPPEDDVES 563
 DB 917 -KEDAKPKKTEVAK--KEPDDAKA 938

RESULT 5

A37221
 neurofilament triplet H protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C/Accession: A37221; A25649; A30796; A32757; B25649
 R:Chin, S.S.M.; Liem, R.K.H.
 J. Neurosci. 10, 3714-3726, 1990
 A>Title: Transfected rat high-molecular-weight neurofilament (NF-H) coassembles with vime
 A/Reference number: A37221; PMID:91038277; PMID:2230956
 A/Accession: A37221
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A/Residues: 1-1072 <CHI>
 A/Cross-references: UNIPROT:Q35482; GB:AF031879; NID:G2642597; PIDN:AA87068.1; PID:G264;
 R:Robinson, P.A.; Wion, D.; Anderton, B.H.
 FEBS Lett. 209, 203-205, 1986
 A>Title: Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H).
 A/Reference number: A25649; PMID:87080760; PMID:2878828
 A/Accession: A25649
 A:Molecule type: mRNA
 A/Residues: 230-318; 472-542 <ROB>
 A/Cross-references: GB:M37227
 R:Bautigny, A.; Pham-Dinh, D.; Rouseel, C.; Felix, J.M.; Nussbaum, J.L.; Jolles, P.
 Biochem. Biophys. Res. Commun. 154, 1099-1106, 1988
 A>Title: The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ dete
 A/Reference number: A30796; PMID:88309090; PMID:2457365
 A/Accession: A30796
 A:Molecule type: mRNA
 A/Residues: 266-421, 'T', 423-427, 'T', 429-542, 'V', 556-566, 'E', 568-613, 'A', 615-725, 'S', 727-7
 A/Cross-references: GB:M21964; NID:G205685; PIDN:AAA41695.1; PID:G205686
 R:Liéberburg, I.; Spitzer, N.; Snyder, S.; Anderson, J.; Goldhaber, D.; Smulowitz, M.; C
 Proc. Natl. Acad. Sci. U.S.A. 86, 2463-2467, 1989
 A>Title: Cloning of a cDNA encoding the rat high molecular weight neurofilament peptide
 A/Reference number: A32757; PMID:89184647; PMID:2928342
 A/Accession: A32757
 A>Status: preliminary
 A:Molecule type: mRNA
 A/Residues: 559-566, 'E', 568-967, 'V', 969-997, 'GST', 1001-1022, 'E', 1024-1072 <LIE>
 A/Cross-references: GB:J04517; NID:G205679; PIDN:AAA41692.1; PID:G205680
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Query Match 6.1%; Score 197.5; DB 1; Length 1072;
 Best Local Similarity 22.1%; Pred. No. 0.035;
 Matches 112; Conservative 82; Mismatches 235; Indels 77; Gaps 20;

QY 11 AAAPAGNPEQRLDYRAAALGPEDEPGAALHPLPRHK---LKEPPLASSQGS 66
 DB 548 AEAKSPAEKSPAEVSPATVKSPEAKSPAEK--SPAETKSPATVKSPEAKSPAEKS 606
 QY 67 PAPSPPGCGGKGLLPPGA-----APGQGESEWGSVPLPCPPATKQIGGE-P 118
 DB 607 PAEVKSPVEAKS-----PEAKSPASVKSPEKASPAEAKSPAEKSPATVKSPEAKSP 661
 QY 119 AAAGACSPKPKYQAVLPLOTGLVAAKEPTPWAGDKGAASPAATASDPAGPPLPLP 178
 DB 662 AEVKSPTVTKSPAEKSPVEVKS--PASVKSPE--AKSPAGAKSPAEKSPVVAKSPAEK 719
 QY 179 GPPPLAATATAGTLAASBGRKSMKSPPLGGGSGGASQAACIKQILLQLDLIEQQOQ 238
 DB 720 SPAEAKPPAEKSPA-----KSPAEKSPAEKSPAEKSPV---EVKSPKAKS 768

Db 3025 -----DSPICAVENTPLPDDSPICAVENTPLPDDSPICGPPHQPVSXTLHNT 3075
 QY 380 EURSQR-----TPKPR-----SSVTPP----- 398
 Db 3076 NLVSPSRSAANVPLPDBSDSGFYAVNIPLPDBPTDDEPFSNOSRAQASVSGS 3135
 QY 399 -----RLSTPQKPSSTH-----PKKAFSSSEIDLPY-----LSTTMYLC-----R 435
 Db 3136 SYKINTGRIPIFAMQAFATSHGRSRNRSTSKPSQSAFYKVPALSTYKIPAVNAGS 3195
 QY 436 WQPP-----SPLPRESPPKEETVAR----- 459
 Db 3196 HNAKSKNEKPKCDTPTVLFGSRNISPSQSTTTANISSTLPQNSTAKSHKVAKTPL 3255
 QY 460 CLMPSEV-AGETSVL-----AVPSMRDHSVEPLRDPNP-----SDLENTL-----D 499
 Db 3256 RVFPSSMPADDIDELVYDQINRAVNTKPLPKSPLOQPPPEYSVTTDYKQVRRPMSE 3315
 QY 500 DSVFSKRAKLEIDEXRRKMDIQIRIQRILQRLRMVKKGIQSESEPTVSFPPEPD 559
 Db 3316 DETIALLINMNDTENDAEPIIDIKSIRAOVLPKOI-----KQANKFVP--LDWMTETE 3367
 QY 560 ---DVSLMITPLPVLVAFGRPLPKLTTPQNFELPMLDERSCRLBIOCK 605
 Db 3368 SAADAGLELSP-----KQPKL-----FWSKSKRLSNINLXDK 3401
 RESULT 8
 QFM5H
 neurofilament triplet H protein - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
 C:Accession: J0368; A43778; S42616
 R:Julien, J.P.; Cole, F.; Beaudet, L.; Sidky, M.; Flavell, D.; Grosveld, F.; Mushinski, Gene 68, 307-314, 1988
 A:Title: Sequence and structure of the mouse gene coding for the largest neurofilament B
 A:Reference number: J0368; MUID:89121513; PMID:3220257
 A:Accession: J0368
 A:Molecule type: DNA
 A:Residues: 1-1087 <JUL>
 A:Cross-references: UNIPROT:P19246; GB:M23349; GB:M24496; NID:g200034; PIDN:AAA9813.1; R:Shmildman, P.S.; Carden, M.J.; Lees, J.F.; Lazzerini, R.A. Brain Res. Mol. Brain Res. 4, 217-231, 1988
 A:Title: The structure of the largest murine neurofilament protein (NF-H) as revealed by
 A:Reference number: A43778
 A:Accession: A43778
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 'M', 1-132, 'QA', 134-199, 'R', 200-280, 'T', 282-491, 'G', 493-533, 'GEAKSP', 534-545, R:Carden, M.J. submitted to the EMBL Data Library, March 1994
 A:Reference number: S42616
 A:Accession: S42616
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-132, 'QA', 134-199, 'R', 200-280, 'T', 282-491, 'G', 493-533, 'GEAKSP', 534-545, 'R', A:Cross-references: EMBL:Z31012; NID:g463249; PIDN:CAA83229.1; PID:g463250
 C:Genetics:
 A:Gene: nfh
 A:introns: 290/1, 356/3; 398/2
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil; cytoskeleton; heterotrimer; intermediate filament; nerve; phosph
 F:1-97/Domain: amino-terminal <NTP>
 F:198-408/Domain: rod hecaton predicted <ROD>
 F:409-1087/Domain: carboxyl-terminal <CTE>
 F:1519-886/Region: 6-residue repeats
 F:520, 526, 538, 544, 550, 556, 562, 568, 574, 580, 586, 592, 598, 604, 610, 616, 622, 628, 634, 640, 64
 73, 885/Binding site: phosphate (Ser) (covalent) #status predicted
 F:772/Binding site: phosphate (Thr) (covalent) #status predicted
 Query Match 5.9%; Score 190; DB 1; Length 1087;
 Best Local Similarity 23.4%; Pred. No. 0.079;

Matches 125; Conservative 52; Mismatches 222; Indels 136; Gaps 25;
 QY 11 AAAPAGNEORLDYERAAALGGPEDEPGAENHFLPRHRK-----LKEGPPPLASSGGS 66
 Db 636 ATYKSGEAKSPSEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKS 694
 QY 67 P-----APSPAGCGGKGRGLLPAGAAPGQOESWGSVPLPCPPATKAGIGIGEP 118
 Db 695 PAEVKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKS 743
 QY 119 AAAGACSPRPKQAVLPQTQSLVAAAKEPTTMADKGAASPAATASDPAPPLPLP 178
 Db 744 GEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKS 797
 QY 179 GPPPLAFTVAGTLAASBGRKMSKSPGCGGGSASQAACLKQILLQLDLICQOQ 238
 Db 798 IKRP-----AEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKS 825
 QY 239 QLOAKEKEIEELKSERDTLLARIERNRMQVKKONERHKLFGYETEEREELSLP 298
 Db 826 PLDVKSPPEAQTPVQEAATVPTDIRPEQ-----VKSAPAEKAK-----SPEKEAKTSE 874
 QY 299 KI---KLECC-PELSTSQTLPPKPPSCGSGGKGRKSPGSTTEKTPVKLAP-EFSK 353
 Db 875 KAAPKKEVKSPEKKEVKAKEPKVE-----EERTLPPTKEAKSKKDBAPKEAPK 927
 QY 354 VKTKTPKSPKPEKPGCSLETVCCKELRSQETPEKPR---SSVTPPLSTPQKGPSTH 410
 Db 928 PKVEEKETP-TEKPDSTREA-----KKEEGEKKAVASSEETPAKGVKE--AK 977
 QY 411 PKEX--AFSSSEIDLPYLSTTMYLCRMHQPPSPPLREBSSPKETVARCLMPSSVAG 468
 Db 978 PKKETTETKTEADT-----KAKESK-PTEKEXKKE-----MP----- 1012
 QY 469 ETSVLAPEKRDHSVPL---RDPNSDLI---NDDSVFSKRAKLEIDEXR 517
 Db 1013 -----AAPEKDKTEKTEKTESRKEPKWEAKVEDKSLSKPEKPKTEKAEK 1062

RESULT 9
 T08599
 probable transcription factor CA150 - human
 C:Species: Homo sapiens (man)
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
 C:Accession: T08599
 R:Sun, C.; Hayashi, T.; Liu, Y.; Lane, W.S.; Young, R.A.; Garcia-Blanco, M.A. Mol. Cell. Biol. 17, 6029-6039, 1997
 A:Title: CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is a
 A:Reference number: Z16449; MUID:97459702; PMID:9315662
 A:Accession: T08599
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1098 <SUN>
 A:Cross-references: UNIPROT:Q14776; EMBL:AF017789; NID:g2460123; PIDN:AA80727.1; PID:g24
 A:Experimental source: cell line HeLa
 C:Genetics:
 A:Gene: CA150
 A:Function:
 A:Description: HIV-1 Tat transcriptional coactivator
 C:Keywords: nucleus; transcription regulation
 F:131-168/Domain: WW repeat homology <WW1>
 F:429-466/Domain: WW repeat homology <WW2>
 F:527-565/Domain: WW repeat homology <WW3>
 Query Match 5.9%; Score 190; DB 2; Length 1098;
 Best Local Similarity 19.8%; Pred. No. 0.08;
 Matches 133; Conservative 80; Mismatches 221; Indels 238; Gaps 30;
 QY 18 NP-EORLDYERAAALGGPEDEPGAENHFLPRHRKKEPPPLASSGGSPPASGCGG 76
 Db 15 NPEELMAQQOQALRFGRGAPAPPVNAVNRGPP-----LMRPPPPGMMRGPPPPPP 65
 QY 77 KGRGLLPAGAAPGQOESWGSVPLPCPPATKAGIGEPAAAGSGSPRPKYQAVLP 136

```
Db 66 -----PGRPP-----FDPNMF-PMPP-----GGIPRPMKPRHLQRPPEH----- 100
Qy 137 IQTGSLVAAAKEPTPWAGDKGAASPAATASDPAGPPLPLPQPPPLATATAGTLAASE 196
Db 101 -----PPPMSS-----SMPPPPGMPPGMPVTAAGTAPLPT-----E 134
Qy 197 GRWKSMTKSPGLGGGGSGAS-----QAAGLKQILLLQLDLLEQQOQOOLA 242
Db 135 EIWVE-NKTPDGKVVYYNARTRESAMTKPDGVKVIQQSELTPMLAAQAVQAQQAQQA 193
Qy 243 KEKEIEELKSERDILLARIEMEREMQLVKDKNEKERHKLFGQYETEEEREELSEKIKL 302
Db 194 QAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 249
Qy 303 ECQPEL--SETSQTLPPKPPSCGSGKGHKRKSFGSTERKTPVKKLADEFKVKTKTEK 360
Db 250 QVQAQAVGASTPTTSSPAP-----AVSTSTSSSP-----SSTTSTTTAT-- 290
Qy 361 HSPITKEEPCGSLSETVCKRELASQETPER-----PRSSVDTPPRLSTP-QKQPSHPE 413
Db 291 -----SVAQTVSTPTTQDQ--TPSSAVSVATPTVSVSTPARKATATVQVPPQHPQT 339
Qy 414 -----KAF-----SSIEDLPY-----LSTTEM 431
Db 340 LPRAVPHSVPOPTTALPAPEPVWVPPFRVPLPGMPTPLPGVAMQIVSCPYKTVATTKT 399
Qy 432 YLCRWHPSPPLPLRESSPKKEETVARGCLMPSSVAGETSVLAVPSWRDHSV----- 483
Db 400 GVLPGMAPPIVPM-----IHPQVAILASPATLAGAT--AVSEWTEYKADGKTYXX 448
Qy 484 -----EPIADRPSPDLLENLDDSVSKHAKLEDEKRR 517
Db 449 NNRTLESTWEKQELKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE 494
Qy 518 KRWDIQRIREQRILOLRMYKKKGIQSEBPVTSFPPEPDVSEIMTTPFLPV----- 572
Db 495 KEKPIKEIEE--PKEEMTEBERKAQAQKAVATAPITFG-----TPMCVWTGDE 542
Qy 573 -VAFGRPLPKLT 583
Db 543 RVFFYNPTRLS 554

RESULT 10
A34327
neurofilament triplet H1 protein - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A43427
R:Sopet, D.R.; Beasley, L.L.; Willard, M.B.
J. Biol. Chem. 267, 17354-17361, 1992
A>Title: Evidence for unequal crossing over in the evolution of the neurofilament polype
A:Reference number: A43427; MUID:92381055; PMID:1512270
A:Accession: A43427
A:Molecule type: DNA
A:Residues: 1-666 <SOP>
A:Cross-references: UNIPROT:Q28687; GB:M94315; NID:g164990; PIDN:AA57152.1; PID:g601930
A:Experimental source: brain
A>Note: sequence extracted from NCBI backbone (NCBIN:112010, NCBIP:112011)
C:Superfamily: neurofilament triplet H protein
C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Query Match 5.9%; Score 189.5; DB 2; Length 606;
Best Local Similarity 20.4%; Pred. No. 0.046;
Matches 114; Conservative 67; Mismatches 250; Indels 127; Gaps 15;

Qy 4 RSAVFAKAAAP-----GNPEQRLDYERAA-----LGGPEDEPGAEEAHFLPRHRK 51
Db 107 KSPVKEEAKSPBAKSPBAKSPBAKSPBAKSPBAKSPBAKSPBAKSPBAKSPBAKSPBAK 163
Qy 52 LKEPGRPLASOGGSPAPSPAPGCGGRGLLPAGAAPQOEESWGSGVPLPPEPPATQ 111
```

```
Db 164 AKSPBAKSPBAKSPVKEBAKSPBAKSPVKEBAKSPBAKSPBAKSPV-----KE 216
Qy 112 AGTGGPAAAGCGCPRPYQAVLPIDTGSVLAALKETPWAGDYGGAASPAATASDPAG 171
Db 217 EAKSPBAKSPBAKSPVKEBAKSP-----EKAESP-----EKEBAKSPBAKSPBAK 264
Qy 172 PPPLPQPPPLAPATAGTLAASGRWKSMTKSPGLGGGGSGASQAACLQIILLQLD 231
Db 265 KSPBAKSPVKEBAKSPBAKSPBAKSPVKEBAKSP----- 299
Qy 233 LIEQOQOOLAQAKEIEELKSERDILLARIEMEREMQLVKDKNEKERHKLFGQYETEE 291
Db 300 --EKAESPVEBAKSPBAKSP-----VBEBAKSPBAKSPVKEBAKSPBAKSPVYK 349
Qy 292 EETELSEKIKLACQPELSETSQTLPPKPPSCGSGKGHKRKSPPGSTERKTPVKKLAPEF 351
Db 350 EBAKSPBAKSPVKEBAKSPBAKSP-----KAKSPV--KEBAKSPBAKSPVYK 397
Qy 352 SKVTKTTPHSPITKEEPCGSLSETVCKRELASQETPERRSSVDTPPRLSTPQKQPSH 411
Db 398 EBAKSPBAKSPVKEBAKSPBAKSPBAKSPBAKSPBAKSPBAKSPBAKSPBAKSPBAK 457
Qy 412 KE-----KAFSSIEDLPYLSTTEMYLICRWHP----- 439
Db 458 KEAPRAVEKPESTAAKEDKDAEDKKAAAPAKMGKEBAKPEKTEVAKKEBEDAKAKE 517
Qy 440 PPSPLPLESSPKKEETVARGCLMPSSVAGETSVLAVPSWRDHSVPLDNPDLLENID 499
Db 518 PPKPTPEKEPEKEKEKEETPA-----APVKEA-----KEEARPEKEPK--TEAKAKED 563
Qy 500 DSVFSKRAHAKLEDEKRR 517
Db 564 DVALSKEPEKPTERAER 581

RESULT 11
T28770
hypothetical protein W03D2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28770
R:Roehlfing, T.; Wohldmann, P.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of C. elegans comid W03D2.
A:Reference number: Z20519
A:Accession: T28770
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-539 <ROH>
A:Cross-references: UNIPROT:Q8MQG9; EMBL:AF000298; PIDN:AAQ48255.1; GSPDB:GN00022; CESP:V
A:Experimental source: strain Bristol N2; clone W03D2
C:Genetics:
A:Gene: CESP:W03D2.1
A:Map position: 4
A:Introns: 40/3; 86/3; 115/3; 146/3; 173/3
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match 5.8%; Score 189; DB 2; Length 539;
Best Local Similarity 24.8%; Pred. No. 0.043;
Matches 109; Conservative 23; Mismatches 132; Indels 176; Gaps 21;

Qy 9 KAAAPAGGNPQRLDYERAAALGGPEDEPGAEEAHFLPRHRKLKBPGLASOGGSPA 68
Db 249 KQOAPPAGSP-----PPPPP-----KGSPPLAGS--GSP 277
Qy 69 PSPAGCGGGRGLLPAGAAPQOEESWGSGVP-----LPCPPATKQAGIGGEPAATAA 123
Db 278 PPPA-----AGSPPPPRT-----GSPPPPTGSPPPPA-----GGSPPPPPA 315
Qy 124 GCSPPRYQAVLPIDTGSVLAALKETPWAGDYGGAASPAATASDPAGPP----- 174
Db 316 GSPPPPPPPRGPSP-PTGSL-----PPPOA-----GGSPPPATGSPSP--PPPRQKQAPRR 363
```

Query Match	5.8%;	Score 188;	DB 2;	Length 2715;
Best Local Similarity	21.4%;	Pred. No. 0.25;		
Matches	125;	Conservative	41;	Mismatches 175; Indels 242; Gaps 26
QY	11	AAAPAGNPEQRDLDYERAAALGGPDEPGAAEAFHFLPHNRKIKE--PGPRLASSQ--GSPS	67	
Db	70	APPPGSGSP-----GGPRGPRPAAVMHNHNLHQQOOOHDPRRHMQOOHHNGSR	117	
QY	68	APSPAGCGGKGGKGLLPAGA---APGQGEESGSGSVPLRSPRPA-----TK	110	
Db	118	APPP-----PGAPRHNAGVKEEY--THLPRRHPRPAAGRYHADBNMDPRRYG	163	
QY	111	QAGIGGE-----PAAAGAGGSP--RPKYQAVLPLQ-----TSLVAAKE	148	
Db	164	QPLRGKGRPQQQQQPHRPPQQRRPQQRRGSGSPNNRPQQRVTPGRRPGGPTPTLNLSSNSP	223	
QY	149	FTP---WAGD-----KGAASPAATASDPAAPRPLPLGPRPLAFTATAGTTLAASBGR	198	
Db	224	PPRRPHRYANTYDPOOAAASAAAAAQQOQGGRRPRGNHGR---PQHPSPRGQOOQG	281	
QY	199	W-----KSMRKSPPL-----GGGGSGGASSQAACLKQTLILQL	230	
Db	282	WAPRRPRYSPQLGSGQQYRTPTPTTSKGGSYPPRAHGONSSGYSSP-----	329	
QY	231	DLIEQOQOQLOAKEKEIEELKSEKSDTLTALIERMERRMQLVKDNEKERHKLFGYEETE	290	
Db	330	---QQOQQOQQOQQOQOQA-----QQQPGRVGGRPRPT	359	
QY	231	REETLSSEKIKLGCQPELSETSOTLP-----PKPFGSGSGGKHK-----RKS	333	
Db	360	GQQPP-----QONTPTPSQYSBYPQRYTPTPGLPRAGGNNHRAATSTHOYREPRBNBW	410	

[illegible]

```

RESULT 14
T51023
hypochemical protein B7F21.40 [imported] - Neurospora crassa
C|Species: Neurospora crassa
C|Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C|Accession: T51023
R|Schulle, U.; Allyn, V.; Hohnsiefel, J.; Brandt, P.; Fattmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A|Reference number: 225286
A|Accession: T51023
A|Status: preliminary
A|Molecule type: DNA

```

A:Residues: 1-2649 <SCH>
A:Cross-references: UNIPROT:Q9P3J0; EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.40
A:Experimental source: BAC clone B7F21; strain OR74A
C:Genetics:
A:Gene: NCSP:B7F21.40
A:Map position: 6
A:Introns: 1619/3; 2584/1

Query Match 5.7%; Score 184.5; DB 2; Length 2649;
Best Local Similarity 23.2%; Pred. No. 0.36;
Matches 118; Conservative 60; Mismatches 171; Indels 159; Gaps 28;

QY 34 PEDDEPAAAHPLPR---HRKL-----KRRGPRLAASGGSPAPSPAGCGKRGILLPA 85
DB 1656 PTPSAGPRRRYDVPSSSGRRPLAABAEPPQ--AKSEAAPRANQPF-----SRFQVTP 1708
QY 86 GAAPGOEESWGSVLP-PCP-PPATKAGIGGEPAAAGAGCSRPYQAVLPQTGSLVA 144
DB 1709 QASPVVHTLVORSPTVMAPLPPA-----PVAAASAAASFG-----TPSRGPANVS 1755
QY 145 AAKEPTWAGDKGAASPAAT--ASDPAGPPPLPLFGPP-----PLAPATAG 190
DB 1756 QTMSPVPHF-----LRQPTATFVTEREG-EPFVGGPQAPQTHQPVRISSQKTAAPVSSS 1809
QY 191 TLASRGKRSKRSPLGGGGSGASSQAACLKQIILLQDLIEQQQQQLQAKETIEL 250
DB 1810 MFSASA-----MRSAGWIPRANRPTPLLSQ-----QHLRVRRI 1846
QY 251 KSERDTLLARIERMERMOVLVKDNEKERHKLFGYETEREEETELSEKIKLECOBELS 310
DB 1847 RGRREGQI-RMELQGRREGRRPRMERPAHFK-----QRE-----QLHND 1886
QY 311 TQQT-LPRPFGSGSGKHKRSKPGSTERTKTPVKLAPEFSKVTKTPKHSPIKEEP 369
DB 1887 PQNFMPQ-----RSMQRAEAAPWG-----RQDPSPSMAF--SAQSYTP--PIQAPV 1933
QY 370 GS-LSRTVCKRELRSQETEKPRSSVDTPRLST-----PQ 404
DB 1934 RNLLSSV-----PQRTTPQUSPAMERPPVPSTQSPMPTSMQEQVSANTSAQVPVPQ 1988
QY 405 KQSTPRKEKAFSSIEDPLYLSTTEMYLCRMHQPPSPPLPLRESPPKKEETVACLMPS 464
DB 1989 APPATPEPKKYSISISILN-----DDNPPAPAP--APKAVNVAS--MER 2030
QY 465 SVAGETSVLAVPSWRDHSVEPLDPNPS 492
DB 2031 AASTSF-----PPQQMSARPPQPPPT 2053

RESULT 15
T00273
hypothetical protein KIA0595 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004
C:Accession: T00273
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:title: Prediction of the coding sequences of unidentified human genes. IX. The complet
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T00273
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-1520 <NAG>
A:Cross-references: UNIPROT:Q9BZES; EMBL:AB011167; NID:g3043713; PIDN:BAA25521.1; PID:g3
A:Experimental source: brain
C:Genetics:
A:Note: KIA0595

Query Match 5.7%; Score 183.5; DB 2; Length 1520;
Best Local Similarity 20.6%; Pred. No. 0.23;
Matches 153; Conservative 64; Mismatches 216; Indels 311; Gaps 34;

QY 13 AAGAGNPEQRLLYERAAALGGPEDEPAAAHPLPRHKKLKERGPPLASGGSPAPS-- 70

DB 681 SPASSPPEPPVVKPVAAS---PTEQVPSQEMLLAR-----PSPVQVSVAVPPLPSM 731
QY 71 ----PAGCGKRGILLPAGAAPGOEESWGSVLP-PCP-PPATKAGIGGEP-----AAA 121
DB 732 SAALPPAGGLGMPSSLF---PPLQPPSLPLSMGVLPDPFTTHAFLPSWPCYPHVS 788
QY 122 GAGCSRPYQAVLPQTGSLVAALAKEPT--PMAGDKGAASPAATASDPAGPPPLPL 178
DB 789 GYPCLEPPF---PTVLVSGTPGAYAVPPPCSVPMW-----PPAPV- 826
QY 179 GPPPLAPTAAGTLAASRGKRSKRSPLGGGGSGASSQAACLKQIILLQDLIEQQQQ 238
DB 827 --SPYSTCTYG-----PLWGPQ----- 843
QY 239 QLAKEKEIEELKSERDTLLARIERMERMOVLVKDNEKERHKLFGYETEREEETELS 298
DB 844 ----- 843
QY 299 KIKLECOPLSTSGTLPPKPF--SCGRS--GKHKRSPPGSTERTKTPVKKLAP--- 349
DB 844 -----PQHAFFWSTVPPPLPPASIGRAVQPKESRGTAGPENVLPISMAPLSL 896
QY 350 -----EFSKVKTKT-----PKH-----SP-IKEBPCS-----LSETVCKR- 379
DB 897 GLPHGAPQTEPTKQVKEVPKVPASPHKHKVSAVLOSPOKALACVSAEGVTVEEPASER 956
QY 380 -----ELRSQETPEKRSVDTPRLSTPQKGPSTHPKE-----XA 415
DB 957 KPQETETREKRPPLPATKAVPTPQSTVTKLPAVHAPALRLSLPLPTRTQSGEDVQA 1016
QY 416 ESSEI---EDL-PYLSTTEMYLCRMHQPPSP--LPLRBS--PKKEETVACLM 463
DB 1017 FISEIGIEASDLSLEQFEKSEAKKECPAPADSLAVGNSGVYDIPQEKPLRLQAP 1076
QY 464 --SVAGETSVLAVPS--WRDHSVEPL--RDP-----NPSDLLENDDSVFSR 506
DB 1077 ELANVAGLTPPATPPHQLMKPLAAVSLAKAKSPASTAOEGTLKEGVTE-----AKH 1129
QY 507 HAKLELD----- 532
DB 1130 PAAVRLQEGVHGPSRVHVGSGDHQCVRSRTPPKKMPALVPEVSGRMVNVKHQDITIK 1189
QY 533 RLQRLMYKKKGIQSESEPVTSFPPEPD--VESLMTTPPLPVVAFGRPLPKLTPQNFELP 590
DB 1190 VLSL-----CPAAPPFCIAASREPLDHTSSQADPSAPCLA--PSSLSP----- 1234
QY 591 WIDERSRCLEIQKQTPHRTCRK 614
DB 1235 ---EASPCRDMDMTKTPPEPSAKQ 1255

Search completed: March 23, 2005, 15:33:48
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 25, 2005, 21:11:08 ; Search time 7487 Seconds
(without alignments)
3973.755 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 3238
Sequence: 1 MTRSAVFKAAPAGAGNP.....RSRCRLRQKQPHRTCRK 614

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xjh
-Q=cg2_1/USFTO.spool/US10054935/rnacat_23032005_144819_26094/app_query.fasta_1.775
-DB=GenBank -QFMT=fasta -SUFFIX=rge -MINMATCH=0.1 -LOOPC=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US10054935 @CGN 1.1 4198 @runat_23032005_144819_26094 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOC=100 -LONGLOC
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenBank: *
1: gb_ba: *
2: gb_hg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	DB ID	Description
1	3115	96.2	2933	10	BC058629 Mus muscu
2	2306	71.2	2087	10	BC055715 Mus muscu
3	2306	71.2	2095	10	BC043039 Mus muscu
4	1757.5	54.3	106773	2	AC119462_3 Continuation (4 of

C	5	1757.5	54.3	196724	10	AL590963
C	6	1581	48.8	183618	2	AC131356
C	7	1456.5	45.0	194780	2	AC068669
C	8	1116	34.5	655	6	CO724830
C	9	1055.5	32.6	615	6	AX401049
C	10	1053.5	32.5	1290	6	AR379777
C	11	1030.5	31.8	1269	6	BC039449
C	12	1020	31.5	1092	11	BV178745
C	13	991	30.6	16993	2	AC148580
C	14	821	25.4	16993	2	CO842718
C	15	819	25.3	3425	9	AK125654
C	16	819	25.3	3425	9	AK125654
C	17	669.5	20.7	2638	9	AK124185
C	18	656.5	20.3	162385	5	AL929039
C	19	537	16.6	2745	9	HSB607332
C	20	416	12.8	208620	2	AC107643
C	21	381.5	11.8	440	6	CO687002
C	22	349	10.8	230913	2	AC119484
C	23	299	9.2	274059	2	AC123367
C	24	283	8.7	146807	5	BX682558
C	25	274.5	8.5	208620	2	AC107643
C	26	259.5	8.0	2742	6	CO730359
C	27	246.5	7.6	2217	5	AB026623
C	28	246	7.6	3712	9	BC073969
C	29	246	7.6	3712	9	AX777372
C	30	246	7.6	3712	9	AF203032
C	31	246	7.6	3929	9	AB020652
C	32	246	7.6	5483	10	AF076623
C	33	245.5	7.6	274676	1	AB017305
C	34	245	7.6	5373	6	AX746189
C	35	243.5	7.5	16832	1	AB070943
C	36	243.5	7.5	300550	1	AP005030
C	37	243	7.5	4680	9	AY648038
C	38	243	7.5	276289	1	AB017306
C	39	242.5	7.5	157788	2	AY118786
C	40	241	7.4	7020	6	CO724545
C	41	240	7.4	7033	9	AF039571
C	42	239	7.4	6123	10	BC060693
C	43	239	7.4	6143	10	BC062893
C	44	239	7.4	280558	1	AB017301
C	45	238	7.4	9551	6	AR076233

ALIGNMENTS

RESULT 1	BC058629	2933 bp	mRNA	linear	ROD 30-JUN-2004
LOCUS	Mus musculus	RIKEN CDNA 4121402D02	gene,	mRNA (CDNA clone MGC:64809	
DEFINITION	IMAGE:6402215), complete cds.				
ACCESSION	BC058629				
VERSION	BC058629.1	GI:35193129			
KEYWORDS	MGC.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 2933)				
AUTHORS	Klauser, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Burow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marins, K., Farmer, A.A., Rubin, G.M., Hong, L., Stappleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.U., Uedon, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huylk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,				

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butcherfield, Y.S., Krzywinski, M.I., Skalek, U., Smalitus, D.E., Scherch, A., Schein, U.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2933)

Strausberg, R.
Direct Submission
Submitted (22-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIND)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr. Thomas L. Casavant.
Web site: <http://genome.uiowa.edu>
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A., Fishler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K., Schetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J., Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIND at: <http://image.lind.gov>
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21311987, location/Qualifiers

1. 2933

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:64809 IMAGE:6402215"
/tissue_type="Mouse, brain 12.5 dpc"
/clone_lib="NIH_BMAP_FCO"
/lab_host="DH10B"
/note="Vector: pYX-ASC"
1. 2933
/gene="4121402D02Rik"
/note="synonym: 2810017P12Rik"
/db_xref="LocusID:74026"
/db_xref="MGI:1921276"
63. 1913
/gene="4121402D02Rik"
/codon_start=1
/product="Riken CDNA 4121402D02"
/protein_id="AAH58629.1"
/db_xref="GI:35193130"
/db_xref="LocusID:74026"
/translaton="MTWSAVFRAAAAPAGGNPQRDLVRYRAALGGPEDESGAAEAH
FLPRHRLKEPRPLASOGSGSPSPAGGKGRGLLPAGAAPQDQSESGAAEAH
LPCPPATKQAGIGGEFVNAAGAGSPPKQAVLPITGSIIVAAAKEFTPMADKRG
AAPPAATADPAGPPLPLPGPPLATATAGTLAASGKWSIRKSLGGGSGSGAS
SOACAKQIDLLDLLEOQOOLQAEKEIEBELKSRDYLARIEMERMOIYKRD
NEKERHLQGYPEREERAELESEKILKEROPCLCETGQALPSKPFSGSGKGRK
TPGNTGRLKTPVKKAEFSGVKTTPKHPDIKEPGGSISETYCKRELKSOETPEK
RSVDTPRLSTPGKGSSTPKKASSEMEDLPYSTETMYLCRMHQPPSPPLPARE
SSPKKETVARCLMPSSVAGETSYLAVPSRHDVSEPLRDPNSDILLENIDDSVFSKL
HAULEDEKRRKRWDIQRIEQRILQRLQIRMYKKGIQSESEPVTSFPFEPDVESEL
LITPLPVPVAFGRPLPLAPQNFELPWLIDERSRCRLIEIQKHTPHTRCRK"

ORIGIN

Alignment Scores:

Pred. No.: 6.56e-85 Length: 2933

Score: 3115.00 Matches: 592
Percent Similarity: 97.73% Conservative: 10
Best Local Similarity: 96.10% Mismatch: 12
Query Match: 96.20% Indels: 2
DB: 10 Gaps: 2

US-10-054-935-2 (1-614) x BC058629 (1-2933)

QY 1 MetThrMetArgSerAlaValPheIlysaAlaAlaProAlaGlyIyaaPProGlu 20
DB 63 ATGACCATGAGATCCCAAGTGTCAAGGCGCGCGCCCTGCGGCAACCCCGAG 122
QY 21 GlnArgLeuAspArgIuArgAlaAlaAlaIleuGlyGlyProGluAspGluProGlyAla 40
DB 123 CAGCGACTGACTACAGAGGCGCTGCGGCGCTGGGCGCGCGAGCAAGTCCGAGG 182
QY 41 AlaGluAlaHisPheLeuProArgHisArgIyLeuIyGluProGlyProProLeuAla 60
DB 183 GCCGAAGCCCATTTCTTCCCGCGCATCTGAAGCTCAAGAGCGCGGCGCTGGCC 242
QY 61 SerSerGlnGlySerProAlaProSerProAlaGlyCys--GlyGlyIyGlyArg 79
DB 243 TCTTCCAGGCGCGAGCCCTCGCCCTCCAGCGGCTGCGCGCGCGCAAGCGCGG 302
QY 80 GlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGluSerTPGlyGlySer 99
DB 363 GTCCCTTGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 422
QY 120 AlaAlaGlyAlaGlyCysSerProArgProIyGlyGlnAlaValLeuProIleGlnThr 139
DB 423 GCAGCGGCGCTGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 482
QY 140 GlySerLeu---ValAlaAlaAlaIyGluProThrProThrAlaGlyAspIyGly 158
DB 483 GGTCTTATCGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 542
QY 159 AlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeuPro 178
DB 543 GCGGCTTCCCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 602
QY 179 GlyProProProLeuAlaProThrAlaAlaAlaGlyThrLeuAlaAlaSerGlnGlyArg 198
DB 603 GGGCCACACACCTCG 662
QY 199 TrpIySerMetArgIySerProLeuGlyIyGlyIyGlySerGlyAlaSerSerGln 218
DB 663 TGGAGAGATTAAGAAAGACCTCTCGGGGATGGCGCGCGCTCGAGGCTTCAGT 722
QY 219 AlaAlaCysLeuIyGlnIleLeuLeuGlnLeuAspLeuIleGlnGlnGlnGln 238
DB 723 GCGCGCTGCTCAACAGATCTCTTCTGCTGCAATTTGACCTCATCAACGACGACG 782
QY 239 GlnLeuGlnAlaIyGlnIyGlnIleGlnIyLeuIySerGlnIyArgAspThrLeuLeu 258
DB 783 CAGTGGAGCGCAAGAGAAAGATAGAGACTGAAGTCCAGAGAGATAGCTCTT 842
QY 259 AlaArgIleGluArgMetGluArgMetGlnLeuValIyIyAspAsnGlnIyGln 278
DB 843 GCTCGAGTTGAAAGTATGGAAGAGCGAGTCCAGCTGTTGAAGGATTAACGAAGAA 902
QY 279 ArgHisIyLeuLeuPheGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGln 298
DB 903 AGGDAAGAGTCTCTCAAGGCTATTAAGCTGAAGAGAGAGAGAGAGAGAGAGAG 962
QY 299 IyAlIyLeuLeuIyCysGlnIyProGluLeuSerGlnIySerGlnIyThrLeuProProIy 318
DB 963 AAAATTAAATTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1022
QY 319 ProPheSerCysGlyArgSerGlyIyGlyHisIyAspIyGlySerProPheIySerThr 338


```

Db      1023 CTTTCATGATGGCCGAGTGGAGAAAGGACACAAAGGAAACCCCATTTGGAAATACA 1082
Qy      339  GUAUgYsthrProValLysLysLeuAlAProGluPheSerLysValLysThr 358
Db      1083 GAAAGAAAGACTCTCTGTTAAAGAGCTGCTCTGATTTTCAAAAGTCAAAACAAACT 1142
Qy      359  ProLysHisSerProIleLysGluGluProCysGlySerLeuSerGluThrValCysLys 378
Db      1143 CTTAAAGACTCTCCATTTAAAGAGAACCTGTGTCTTCATATCAGAAAGCTTTGTAA 1202
Qy      379  ArgGluLeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProPro 398
Db      1203 CGTGAATTGAGAGACCAAGAAACCCCAAGAAAGCCCGCTTTCAGTGGATACCCACCA 1262
Qy      399  ArgLeuSerThrProGlnLysGluProSerThrHisProLysGluLysAlaPheSerSer 418
Db      1263 AGACTCTGACTCCCAAAAGGAGCCAGACCCCAAGAGAAAGCCTTCTCAAGT 1322
Qy      419  GUAUleGluAspLeuProTyrLeuSerThrThrGluMetTyrLeuGlyCysArgTyrHisGln 438
Db      1323 GAGATGAGAAATTTGGCGTACCTTTCACACACAGAAATGATTTGTGTGCTGGCACACAG 1382
Qy      439  ProProProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrValAla 458
Db      1383 CTTCCCATCATCACCGTACCATTTACGGGAATCTCTCAAGAAAGAGAGACATGTAGCA 1442
Qy      459  ArgCysLeuMetProSerSerValAlaGluThrSerValLeuAlaValProSerTyr 478
Db      1443 AGGAGTGTAGCCATCAGTATGTCAGAGAGAACTTCAGCTTGGCTGTCCTTGG 1502
Qy      479  ArgAspHisSerValGluProLeuArgAspProAsnProSerAspLeuLeuGluAsnLeu 498
Db      1503 AGGAGACACTCAGTAAAGAGCTCTTAAGGAGCCCAATCTTCAGACATTTGGAGAACCTG 1562
Qy      499  AspAspSerValPheSerLysArgHisAlaLysLeuGluLeuAspGluLysArgLys 518
Db      1563 GATGACAGTGTATTTTCAAGAGGATGCAAAACTGGAGCTAGACGAAAGAGAGAAA 1622
Qy      519  ArgTTPAspT1leGlnArgL1eArgGluGlnArgL1leuGlnArgLeuGlnLeuArgMet 538
Db      1623 CGATGGATATTCAGAGATTCAGAGAACCAAGAAATTTTACAGCAGCTGACGTAGATG 1682
Qy      539  TyrLysLysLysGlyL1leGlnGluSerGluProGluValThrSerPhePheProGluPro 558
Db      1683 TATTAAGAAAGAAAGAAATTCAGAAATCTGAGCTGAGCTTACCTTATTTCCCTGAGCCA 1742
Qy      559  AspAspValGluSerLeuMetL1eThrProPheLeuProValAlaAlaPheG1YArgPro 578
Db      1743 GATGATGTTGAAGTTGTGATTAACCCCTTCTTGGCTGTGTAGCATTTGGACGCGCA 1802
Qy      579  LeuProLysLeuThrProGlnAsnPheGluLeuProTTPleuAspGluArgSerArgCys 598
Db      1803 TTGCGAAATTAAGCTTCACAGAACTTGAAGTACCTCGGTGGTAGAGCGAAAGCCGTTC 1862
Qy      599  ArgLeuGluL1leGlnLysGlnThrProHisArgThrCysArgLys 614
Db      1863 AGGTGGAGATCCAGAAAGACACACCTTCAACCGAGCTGTAGGAAG 1910

```

```

RESULT 2
BC055715 2087 bp mRNA linear ROD 21-OCT-2003
LOCUS Mus musculus RIKEN cDNA 4121402D02 gene, mRNA (cDNA clone MGC:66782
DEFINITION IMAGE:6414174), complete cds.
ACCESSION BC055715
VERSION BC055715.1 GI:33585901
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2087)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
22388257
12477932
2 (bases 1 to 2087)
Strausberg,R.
Direct Submission
Submitted (01-AUG-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Bueltow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,L., Marusik,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Frange,S., Kaha,S.S., Loquellano,N.A., Peters,G.J.,
Abramsen,R.D., Muijaby,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Raney,J., Helton,E., Kettaman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shechenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalske,U., Smallos,D.E.,
Schmerch,A., Schein,J.E., Jones,S.U. and Marra,M.A.,
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2087)
Strausberg,R.
Direct Submission
Submitted (01-AUG-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAC Plate: 125 Row: m Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21311987.
Location/Qualifiers
1. 2087
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:66782 IMAGE:6414174"
/tissue_type="Brain, pool of enriched mouse brain tissues
from 13.5,14.5,16.5 and 17.5 dpc"
/clone_1lb="with BMAP_PRO"
/lab_host="DH10B"
/notes="Vector: pIX-ASC"
1. 2087
/gene="4121402D02R1k"
/notes="synonyms: 4930463P05R1k, 2810017P12R1k"
/db_xref="LOCUSID:74026"
/db_xref="WGI:1921276"
15. 1406
/gene="4121402D02R1k"
/codon_start=1
/product="4121402D02R1k protein"
/db_xref="WGI:33585902"
/db_xref="LOCUSID:74026"
/db_xref="WGI:1921276"
/translation="MTWRSVAFKAAAPAGNPFQRUDYERAAALGGPEDESGAAEAH

```

ORIGIN

```

FLRRHKKKEPPLASSQGSPPSPAGCGGKGRGILLPAGARPQOESEWSGVP
LECPPATKQAGIGSEPPVAGAGCSPRPRTQYANLPQTQSIIVNAAKEPTPMAGDGC
AAPPATASDPHPPLPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLP
SQAACIKQILLQLDLIEQOQOLQAKKEKIEBKSEBRTLLAIRMRQOLVYRD
NEKERHKLQGYEPEREBEALSEKIKLEQPELCTESQALPEKPPSCRSQSGHXRK
TFPGNTERKTPYKLAPEPESKVTTPKPSPIKEBPQGSISEVVCSEKRELSQETPKP
RSSVDTPPLSTPQKGPSTHPKFKAFSSMEDPLPYLSTTMTYLCRNHQPPLPLPARE
SSPKKEEYVASKA"

```

Alignment Scores:

Pred. No.:	6, 25e-61	Length:	2087
Score:	2306.00	Matches:	440
Percent Similarity:	97.39%	Conservative:	8
Best Local Similarity:	95.65%	Mismatches:	10
Query Match:	71.22%	Indels:	2
DB:	10	Gaps:	2

US-10-054-935-2 (1-614) x BC055715 (1-2087)

```

QY 1 MetThMetaArgSerAlaValPheLyAlaAlaAlaProAlaGlyGlyAenProGlu 20
DB 15 ATGACCATGATGATCCCGACAGTTCAAAGCGCGCGCGCCCTCGCGCGCAACCCGAG 74
QY 21 GlnArgLeuAspThrGluArgAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
DB 75 CAGCGCATGACTACAGAGCGGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCG 134
QY 41 AlaGluAlaHisPheLeuProArgHisArgLyAsLeuGlyGlyProGlyProProLeuAla 60
DB 135 GCGGAAGCCCATTTCTCTCCCGCGCATCGTACAGTCAAGAGCGCGCGCGCGCGCGCG 194
QY 61 SerSerGlnGlyGlySerProAlaProSerProAlaGlyCys---GlyGlyLySGLYArg 79
DB 195 TCTTCCAGGCGCGGAGCCCTCGCCCTCTCCAGCGCGCTGCGCGCGCGCGCGCGCG 254
QY 80 GlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGln 99
DB 255 GGGTTTATCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 314
QY 100 ValProLeuProCysProProProAlaThrLyGlnAlaGlyGlyGlyGlyGlyProAla 119
DB 315 GTGCCCTTGCGCTGTCCGCCCGACGCTACCAACAGCCCGCATCGCGCGCGCGCGCG 374
QY 120 AlaAlaGlyAlaGlyCysSerProArgProLySGLYAlaAlaValLeuProIleGlnThr 139
DB 375 GCAGCGCGCGCTGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
QY 140 GlySerLeu---ValAlaAlaAlaAlaGlyGlnProThrProThrAlaGlyAspLySGLY 158
DB 435 GGCTCTATCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 494
QY 159 AlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeuPro 178
DB 495 GCGGCTTCCCGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 554
QY 179 GlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlnGlyArg 198
DB 555 GGGCGCACACCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 614
QY 199 ThrLySerMetCargLySerProLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGln 218
DB 615 TGGAAAGATATGAGAAAGACCCCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 674
QY 219 AlaAlaCysLeuArgGlnIleLeuLeuGlnIleuAspLeuIleGlnGlnGlnGln 238
DB 675 GCCGCTGCTCAAAAGATCTCTTCTGCTCAATTGAACTTCATCGAACGCGCGCGAG 734
QY 239 GlnLeuGlnAlaLySGLYGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 258
DB 735 CATTTGACGCGCGCAAGAGAGATATGAGAGAGCTGAAGTCCGAGAGAGATACGCTCTT 794
QY 259 AlaArgIleGluArgMetGluArgMetGlnLeuValLySGLYAspAsnGlnLySGLu 278

```

```

DB 795 GCTCGATTGAAAGATATGAGAGAGCGCGAGCTGCAAGCTGTAAAGAGATTAACGAGAGAA 854
QY 279 ArgHisLyLeuPheGlnGlyGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 298
DB 855 AGGCAACAAGCTGCTCAGAGGCTATGAACTCGAAGAGAGAGAGAGAGAGAGAGAGAGAG 914
QY 299 LySILLeuLeuGlnCysGlnProGluLeuSerGluThrSerGlnThrLeuProProLyS 318
DB 915 AAAATTAAATTTGAGCGCGAGCGCGAGCTTTGCGAAGCTCCCGAGCTGCTTCCAG 974
QY 319 ProPheSerCysGlyArgSerGlyLySGLYHisLySGLYArgLySerProPheGlySerThr 338
DB 975 CTTTCTTCAATGTCGCGCGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1034
QY 339 GlnArgLySGLYThrProValLySGLYLeuAlaProGluPheSerLyValLyThrLySGL 358
DB 1035 GAAAGAAAGACTCTGTATAAAAGCTGCGCTCGAATTTTCAAAAGCTCAAAACAAAGAA 1094
QY 359 ProLyHisSerProIleLySGLYGlnGlnProCysGlySerLeuSerGlnThrValCysLyS 378
DB 1095 CTTAACACTCTTCCCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1154
QY 379 ArgGluLeuArgSerGlnGlnThrProGluLySProArgSerSerValAspThrProPro 398
DB 1155 CTTGAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1214
QY 399 ArgLeuSerThrProGlnLySGLYProSerThrHisProLySGLYAlaPheSerSer 418
DB 1215 AGACTCTCGAGCTCCCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
QY 419 GlnIleGlnAspLeuProLySGLYLeuSerThrThrGlnMetCysArgGlyGln 438
DB 1275 GAAATGAAAGATTTGGCGTACCTTTCCACACAGAAATGATTTGCTGTGGACCGAG 1334
QY 439 ProProProSerProLeuProLeuArgGluSerSerProLySGLYGlnGlnThrValAla 458
DB 1335 CTTCCCGCATCACGTTACATTACGAGAAATCTCTTCCAAAGAGAGAGAGAGAGAG 1394

RESULT 3
BC043039 2095 bp mRNA linear ROD 21-OCT-2003
LOCUS BC043039
DEFINITION Mus musculus RIKEN cDNA 4121402D02 gene, mRNA (cDNA clone MGC:57934
IMAGE:5701204), complete cds.
ACCESSION BC043039
VERSION BC043039.1 GI:27696670
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 2095)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L.H., Shenmen, C.F., Schlier, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, D., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheltz, T.E., Brownstein, M.J., Umed, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loggellano, N.A., Peters, G.J.,
Abramson, R.D., Muljany, S.J., Bosak, S.A., McEwan, P.O.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Pahay, J., Helton, E., Kerteman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Boutard, G.G., Bleskesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E.,
Scherer, A., Schein, J.E., Jones, S.J., Jones, S.J., and Marra, M.A.,
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

```

MEDLINE	22388257
PUBMED	12477932
REFERENCE	2 (bases 1 to 2055)
AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (09-JAN-20)

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 213111987.

FEATURES
SOURCE

US-10-054-935-2 (1-614) X BC043039 (1-2095)

QY 359 ProlyHisSerProIleuysGluGluProCyGlySerIeuSerGluThrValaCyGlys 378
Db 1121 CTTAAAGACTCTCCCATTAAGAGAAACCTGTGGTTCATATCAGAACTGTTTGTAA 1180
QY 379 ArgGluLeuArgSerGlnGluThrProGluLeuProArgSerSerValaAspThrProPro 398
Db 1181 CGTGANTTAGAGACCAAGAAACCCAGAAAGCCCGGTCTTCAGTGAATACCCACCA 1240
QY 399 ArgLeuSerThrProGluysGlyProSerThrHisProlysGluLeuAlaPheSerSer 418
Db 1241 AGACTCTCGACTCCCAAAAGGAGCCAGACCCACCAAGAGAAAGCCTTCTCAAGT 1300
QY 419 GluIleGluAspLeuProTyrIleuSerThrThrgluMetTyrIleuCyAspGTPHsGln 438
Db 1301 GAGATGGAAGATTGCGGTACCTTTCACACAGAAATGTAATTGTGCTGTGGCACAG 1360
QY 439 ProProProSerProLeuProLeuArgGluSerSerProIlyslsGluGluThrValaAla 458
Db 1361 CTTCCCCATCACCGTTACCATTAACGGAAATCTCTCCAAAGAGAGAGACTGTAGCA 1420

RESULT 4
AC119462.3

WPCOMMENT

Sequence split into 4 fragments LOCUS AC119462 Accession AC119462

Fragment Name	Begin	End
AC119462_0	1	110000
AC119462_1	100001	210000
AC119462_2	200001	310000
AC119462_3	300001	406773

Continuation (4 of 4) of AC119462 from base 300001 (AC119462 Rattus norvegicus clone CH2)

Alignment Scores:

Pred. No.:	3.83e-43	Length:	106773
Score:	1757.50	Matches:	445
Percent Similarity:	30.01%	Conservative:	9
Best Local Similarity:	29.41%	Mismatches:	17
Query Match:	54.28%	Indels:	1043
DB:	2	Gaps:	5

US-10-054-935-2 (1-614) x AC119462_3 (1-106773)

QY 1 MetThrMetArgSerAlaValaPheIleValaAlaAlaProIleArgIlyAsnProGlu 20
Db 24446 ATAAACCAATGAGATCCGAGATGTTCAAGCGAGCGGCGCCCTGCGGCGCAACCCCGAG 24505
QY 21 GlnArgLeuAspTyrGluArgAlaAlaAlaIleuGlyGlyProGluAspGluProGlyAla 40
Db 24506 CAGCGACTGGACTACGAGCGGGCTGCGGCTGTGGCGGGGCCCGAGACGAGTCCGGGGCG 24565
QY 41 AlaGluAlaHisPheIleuProArgHisArgIlyIleuIysGluProGlyProProLeuAla 60
Db 24566 GCCGAAACCCACTTCTCCCGGCAATCGTAAGCTCAAGAGCCCGGGCCCGCTGGCC 24625
QY 61 SerSerGlnIlyIysSerProAlaProSerProAlaIlyCys---GlyGlyIysArg 79
Db 24626 TCTAACCGAGGCGGAGCCCGCGCTCTCCAGCCGCGGCGGCGGCGCAAGGGCGG 24685
QY 80 GlyLeuIleuLeuProAlaGlyAlaAlaProGlyGlnGlnIleuIysSerTrpGlyIlySer 99
Db 24686 GGTTTGTTACTCCCGCGCGGGCGCGCCCGGCGAGAGAGAGAGACTGGGGCGGTTCCG 24745
QY 100 ValProIleuProCyAspProProProAlaThrIlysglnIleuIlyIleuGlyIlyProAla 119
Db 24746 GTCCCTCTGCTGTCCGCCCGCGGTACCAACACCGGCATCGCGGGAGCCAGTCC 24805
QY 120 AlaAlaGlyAlaGlyCySerSerProArgProIlyIlyIleuAlaValaLeuProIleuIthr 139
Db 24806 GCGAGCGCGCGCTGTGCGACCCCGCGCCCAAGATATAGCGGTGCTGCCATTCAAGCG 24865
QY 140 GlySerIeuValaAlaAlaIlysgluProThrProIlyIlyIleuAlaIlyIlyGlyAla 159
Db 24866 GGCTCTCTGTGGCGCGGCCCAAGAGCTACGCCCTGTGGGTGGAGCAAGGGTGGGGCG 24925

QY 160 AlaSerProAlaAlaThrAlaSerAspProAlaIlyIlyProProProIleuProIleuProGly 179
Db 24926 GCTCCCCAGCTGCCACCGCCTCGAGCCGCGGGAGCCCGCCACACTACTCTGCCCCG 24985
QY 180 ProProProLeuAlaProThrAlaThrAlaIlyIlyIleuAlaAlaSerGluIlyArgTrp 199
Db 24986 CCGCCACCCCTCGCGGCCACCGCACTGCCGGAGCCCTGGCCGCGGTAGAGGCAATGG 25045
QY 200 LysSerMetArgIlySerSerProIleuGlyIlyIlyIlySerGlyAlaSerSerGlnAla 219
Db 25046 AAGAGATATAGAGAAAGAGCCCTCTCGGGGGTGGCGGCGCTCGGAGACCTCCAGTCAAGCC 25105
QY 220 AlaCyLeuIlysglnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 239
Db 25106 GCCTGCTCAACACAGATCCTTCTGCTGCAATTGGAATCTCATCGAACAGACGACGACAG 25165
QY 240 IeuGAlaIlysgluIlysgluIlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 257
Db 25166 TTGCAGCCCAAGAGAAAGAGATCGAGAGCTGAAATCGAGAGAGATACGGT-ACGGGA 25224
QY 257 ----- 257
Db 25225 GGGGTTAATCTGCATTAGGGATGGGAGCGAGCTTCGCGATGCTCGGGGAAAGGGGCT 25284
QY 257 ----- 257
Db 25285 GTAGAGGAGAGAGTCCCGCTAGGTTAGGTTAGAGAAATGACCATCATGTAAAGAG 25344
QY 257 ----- 257
Db 25345 TACCTTAGATGTGGGTGTTAGAGACCTGGCATCACTGATACGTTGGTCCGAAAGAA 25404
QY 257 ----- 257
Db 25405 GGTAAAGGCAACCTTGT 25464
QY 257 ----- 257
Db 25465 TGGCAAGGGTTAATTAAATGACAGCCCATTACTTGTGGAAAGCGTTTAGATGG 25524
QY 257 ----- 257
Db 25525 AAAAGACTGATTTACGAGGGGTTAAATGATAGGTGGGTACAGATGTGCTGTATGG 25584
QY 257 ----- 257
Db 25585 AAGGGCTATATTAAAGCGCTGTGATTGGCAGTGCACAGTAGTGAGAGACATAACAG 25644
QY 257 ----- 257
Db 25645 TGTTCCTTAACACCGCTGCAGCATCTTTAGACAGAGGGTTGTACCGGTCTTCGGT 25704
QY 257 ----- 257
Db 25705 CGGTAGGGGTAAGATGCTTTGACAGACTCCTATTATTAGCTTTCTTAACGAGAAACA 25764
QY 257 ----- 257
Db 25765 CTGCTTGTGTTCAAGTTTGTAGTGGGAATCTGAGTTATTAACACACACCACTGAC 25824
QY 257 ----- 257
Db 25825 CAAGTCTTGAAGTACGTGCAGATTTTGAATATCAGCTCATTCAGATATTTCTGTGA 25884
QY 257 ----- 257
Db 25885 ATCCCTCACTCCCAACCTGTTACAGAACTTGAAGGTGAGAGAAAGTGTGTTCT 25944
QY 257 ----- 257
Db 25945 TTCTTTGTACTTGTGGGAAATGACTATAGACTTGTGTTGCACTGCAGACAGACAGC 26004
QY 257 ----- 257

Db 26005 AGCAGCAGCAGCAGCAGCAGCAGCTGCAGTCTTAGGCATTCATTGTTAAACAC 26064
QY 257 ----- 257
Db 26065 ACTCTCACATCAGGAGTGTGGCTGGCCGGGGTGGTGGTATTACAGGATTAACACAC 26124
QY 257 ----- 257
Db 26125 CTTTACCTCCCTAAATAAATACATGGTTTCCACAGCCTGTGTGCTCAGGCTGCCCTC 26184
QY 257 ----- 257
Db 26185 CTTTACCTGTCTCTCCTGAATCTCTCACTGGGGCTGTGTGCACACAGAGCTGAGCA 26244
QY 257 ----- 257
Db 26245 CGGAGCAGAAGAGCTTTCATTTTGTCTGCTTTCTAAATGAGTCCCATCAGGCGCTA 26304
QY 257 ----- 257
Db 26305 TGTCAATGAGAGCTGATCTTGCAGAGGTGAGAGAGAGAGAGAGAGTCTCGTTGAAT 26364
QY 257 ----- 257
Db 26365 GCCCAGGTGTTGATTGTGAGAGTTCTCAACATGAGAGTGTACTGCTCAGTACTGTAC 26424
QY 257 ----- 257
Db 26425 TGCAGCTAGGGTGTGATGAGCACTGTATTACAACAACATATCTTCTTGTGATGGA 26484
QY 257 ----- 257
Db 26485 AGGGGGTGGGAAAAGAGACCAAGTAAGAGAAAGAAATGAAGACTGTGTCAACCAAAAC 26544
QY 257 ----- 257
Db 26545 ATGGTTAAGTAATCTAAGGCTGTGAATCAGAGCAAGATGAAGTGTGCTGACAGGG 26604
QY 257 ----- 257
Db 26605 ATACTGCTACTCTGAGGGTTTTGTCTGCCCTCCCAAGTGATGCCACTCAGAGGCTTT 26664
QY 257 ----- 257
Db 26665 TCAAGTAGGAAAAAGTGGGGTGAAGAGATAGAGCATCCGCTGAGCTCGGCTCAC 26724
QY 257 ----- 257
Db 26725 TGCTGCTGTGTACTAGAGTCTTAAGCTTGAAGCCTGCGCTCAGTCTGTGGC 26784
QY 257 ----- 257
Db 26785 TCTGCTGTAATCTATGCAATGATATATACAGAGCAGCAGTATGATGACCAAGCT 26844
QY 257 ----- 257
Db 26845 CAATTCCTGAGCATGTGTGACGAGCGCAGTTGTCAAGCTTTTGTCAAAACAGTTAATAC 26904
QY 257 ----- 257
Db 26905 CAAGTCTGTATGACAGTACCTTCATTTTGTGATTTTGTGTCTCTTAATTAAG 26964
QY 257 ----- 257
Db 26965 AATTAAATATCTAATATCTAGATAAGTATCTCTCATTAATCCAAATACAAAACAG 27024
QY 257 ----- 257
Db 27025 AATATATGTGCATATGTTTTGGGATTATATTAAGAGTGTGATTAATTAATTAATTAAT 27084
QY 257 ----- 257

Db 27085 TAGTATGTTTAAAGTACTCCCTCAAAATACGACACATATACATTTGAAATGTACTG 27144
QY 257 ----- 257
Db 27145 CTATCCCTAGGACACTGTGTGAGCTCTTGAATCTCTTCTTCACTGTGTGTTAAACATTC 27204
QY 257 ----- 257
Db 27205 CACTGAGTCTAGATTTTGGATGAACAAGAGCAGATAGAAATTTGGACAACGTGGGAC 27264
QY 257 ----- 257
Db 27265 ACCTTAAAGACACAGGCTGAGATATGATCTATGTGTGTAGAGTCTTGCCAGC 27324
QY 257 ----- 257
Db 27325 ACAAGTGGAGGCGCTGTGTCTGTCTCTAGACACACATACAAAGTAGAAGACATAGAG 27384
QY 257 ----- 257
Db 27385 AGGTTCAGTTGAGTGGCTCCAAAAGACTTGAGCCACATGGCTAGAGAGGCT 27444
QY 257 ----- 257
Db 27445 TTGCCATTCATTAAGGACCTGTGACCCGCCAGTGGGAGAGATCTTTCAGAGA 27504
QY 257 ----- 257
Db 27505 ATGATCTGGATTTGCCATCCCTCTTACCATTCAGACTAATCTTATTAATTCATTGTC 27564
QY 257 ----- 257
Db 27565 TTTTGCTTACTGGTTTAAAGGAAAGCTATTCAGTGTGCTGCGCAATTCAAGAGAAA 27624
QY 257 ----- 257
Db 27625 GTTGAATTTGAGGATTCACACAGCAACATGTAGAGCGCAATTCATGAGGGTTAAGT 27684
QY 258 ----- 259
Db 27685 ACTGTGTCTAATTTCTATGTATTAAGTCTGCAATCTGTGACTTCTTTAGCTCTTGCT 27744
QY 260 Arg11egl1uargmetgluargatmeGlnleuVal1lylelysaPhearg1ulysgluArg 279
Db 2745 CGGATTGAACGTATGGAAGAGCCGAGTGCAGCTGTAAAGAGGATACGAGAGAGAAAG 27804
QY 280 His1yLeuPheglng1yTg1uthr1glu1uarg1uGlu1uthr1gluLeuSer1uly 299
Db 27805 CACAGCTGTTTTCAGGGCTATGAACTTGAAAGAGAGAGAGAAACGGAGTGTCTGAGAAA 27864
QY 300 Ile1yLeuGlu1uCyser1uProGluLeuSer1uThrSer1uIle1uPro1uPro1uPro 319
Db 27865 AATTAACTGAGAGGCCAGCCGAGCTCTGCGAAGCATCCAGGCTCTGCTTCAAGCT 27924
QY 320 PheSerCyser1yArgSer1yLyser1yIle1y 330
Db 27925 TTCTCATGTGCGCGAGTGGCAAGGACACAAAAGGTGTGTGTGACTTATTTATGA 27984
QY 330 ----- 330
Db 27985 ACTGAAAAGATAGAGAACTGTAGTGTGATGAAATATGCGCTTGATCTGTGATG 28044
QY 330 ----- 330
Db 28045 GCTTTAGAGCAGCCATAGCCTGAATTGAGAAGTCTGTTTATCTGAGCATGTGGCTTCC 28104
QY 330 ----- 330
Db 28105 CCTGTAATCAGAGACTCAGTTGAGGGCAGCGCTGAGATAGAGAACCTGTTTTTACTT 28164
QY 330 ----- 330
Db 28165 GTCTTGAGCATTTTTTTTGTATACAGTGTCTTAAGTACAGAAATGACAGATATATTTTAAA 28224

```

QY      330 ----- 330
Db      28225 GAAATGATAGGGGTTGGGATTTAGCTAGTGTAAGCGCTTGCCTAGCAAGCTC 28284
QY      330 ----- 330
Db      28285 AAGCCCTGGGTTCCGTCCTCCAGCTCCGAAAAAAGAACCAAAAAAAGAGATA 28344
QY      330 ----- 330
Db      28345 GTCTTAAGATAGACGCTTGCTGATTGACCAAGGCTTTTCTGTACTTTAAGCATG 28404
QY      330 ----- 330
Db      28405 ACACTACTGTCATGAGCTTCCAAATGAACAGATAGATTATTAAGTACCTGCTACTC 28464
QY      330 ----- 330
Db      28465 TTGCCTCAAGTATATTCAAAATCAAACTTATTGAAATTAATTATTTTATTAACATTT 28524
QY      331 ----- 345
Db      28525 TCTTTGCTACTATATAGAAAAACCCATTTGGAATACAGAAAAAGAACTCCGTTTAA 28584
QY      346 LysLeuAlaProGluPheSerLysValLysThrLysThrProLysSerProLysLys 365
Db      28585 AAGCTGCTCTCTGAATTTTCAAAATTCMAAAACAAACCTCTAAGCATCTCCCATTTAA 28644
QY      366 GluGluProCysGlySerLeuSerGluThrValCysLysArgLysLeuArgSerGlu 385
Db      28645 GAGGAGCCCTGTGTTTCATATCAGAAACTGTTTGAACGTGATTTAGAGAGCCAGAA 28704
QY      386 ThrProGluLysProLysSerValAspThrProProLysLeuSerThrProGlnLys 405
Db      28705 ACCCGAAGAAAGCCCGGTCTTCAGTGATACCCACCAAGACTCTGATCCCAAAAG 28764
QY      406 GlyProSerThrHisProLysGluLysAlaPheSerSerGluLeuAspLeuProTyr 425
Db      28765 GGACCCAGCAGCCACCCCAAGGAGCAAGCTTCTCAAGTATGAGATGTAAGTTTGCCCTAC 28824
QY      426 LeuSerThrThrGluMetTyrLeuCysArgTyrHisGlnProProLysProLysPro 445
Db      28825 CTTTCCACCAAGAAATGATTTGTGCTGGACCAAGCTCCCATCAACGGTTTACA 28884
QY      446 LeuArgLysSerSerProLysLysGluGluThrValAlaArgCysLeuMetProSerSer 465
Db      28885 TTACGGGAATCCTCTCCAAAGAAAGAGAGACGTGTACA-----AGT 28926
QY      466 ValAlaGlyGluThrSerValLeuAlaValProSerTyr 478
Db      28927 AAGGCATAGCGAACACTTGCTCTT---ATACCTAGTGG 28962

RESULT 5
AL590963/c 196724 bp DNA linear ROD 05-APR-2002
LOCUS      Mouse DNA sequence from clone RP23-395E10 on chromosome 11,
DEFINITION complete sequence.
ACCESSION AL590963
VERSION   AL590963.11 GI:20068458
KEYWORDS  HTG.
SOURCE     Mus musculus (house mouse)
ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS   Lovell, J.
TITLE     Direct Submission
JOURNAL   Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humphrey@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Apr 7, 2002 this sequence version replaced gi:16504739.
COMMENT   During sequence assembly data is compared from overlapping clones.

```

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-395E10 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

FEATURES

```

source
    location/Qualifiers
    1..196724
       /organism="Mus musculus"
       /mol_type="genomic DNA"
       /db_xref="taxon:10090"
       /chromosome="11"
       /clone="RP23-395E10"
       /clone_11b="RPI-23"

```

ORIGIN

```

Alignment Scores:
Pred. No.:      6,41e-43      Length:      196724
Score:          1757.50      Matches:      440
Percent Similarity: 29.53%      Conservative: 8
Best Local Similarity: 29.00%      Mismatches: 10
Query Match:    54.28%      Indels:      1059
DB:              10      Gaps:      4

```

US-10-054-935-2 (1-614) x AL590963 (1-196724)

```

QY      1 MetThrMetArgSerLeuValPheLysAlaAlaAlaProAlaGlyAsnProGlu 20
Db      62746 ATACCATGATGATCCGACAGTTCAGAGCGCGCCCTGCTCCGCGCAACCCCGAG 62687
QY      21 GlnArgLeuAspTyrGluArgAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
Db      62686 CAGCGACTGAGACTACAGAGCGGCTGGCGGCTGGCGGCCGAGACAGTCCGGGGCG 62627
QY      41 AlaGluAlaHisPheLeuProArgHisArgLysLeuLysGluProGlyProLeuAla 60
Db      62626 GCCGAAGCCCATTTCTCCCGCGGCATCTGATCAAGAGCCGGGGGCCCGCTGGCC 62567
QY      61 SerSerGlnGlyGlySerProAlaProSerProAlaGlyCys---GlyGlyLysGlyArg 79
Db      62566 TCTTCCAGGGGGGGAGCCCTTCCCTTCCAGCGGCTGGCGCGCGCAAGGGCGGG 62507
QY      80 GlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGluLeuSerTyrGlySer 99
Db      62506 GGGCTTTTACTCCACACCGGGGGGGCCCGCGGCAACAGAAAGAGCTGGGCGGTTCG 62447
QY      100 ValProLeuProCysArgProProAlaTyrLysGlnAlaGlyTyrLeuGlyGluProAla 119
Db      62446 GTCCCTTCCCTCGTCCGCCGCCACGACTAAACAAGCGGCGATCGCGGGAGCAAGTC 62387
QY      120 AlaAlaGlyAlaGlyCysSerProArgProLysTyrGlnAlaValLeuProLysGlnThr 139
Db      62386 GCAGCGGGCGTGGCTGCAGCGCCCGCCCAATATACAGCGGTGTGCTGCATTGCAGCG 62327
QY      140 GlySerLeu---ValAlaAlaAlaLysGluProThrProTyrAlaGlyAspLysGlyLys 158
Db      62326 GGCTTCTATCTGTGTGGCGGCGCCCAAGAGCTTACGCCCTTGGGCTGGGAGCAAGGGTGG 62267

```


Db 60106 AATAGCAATCATTAATTTGATGTAAGTACTGATCCCTAGAGGACGCTGTGAGCC 60047
QY 256 ----- 256
Db 60046 CTGATCTTCTGTCATGTGCTGTTTAAACATTCACCTGTGCTAGAGTTTGGAGTAA 59987
QY 256 ----- 256
Db 59986 CATAGACAGATAGAAATTTGGACAAACCGTAGACACTTAAAGAACACAGGCGTAGAA 59927
QY 256 ----- 256
Db 59926 TGTAGCTCAGTGGTGTGATGCTTAGAGTGTGGCCAGGACGCTCAGAGGCCCTGATT 59867
QY 256 ----- 256
Db 59866 CCATCCCTAGACACACAGACGACGACGACGAAACAAAGTTAAATTAGGGGGGTCC 59807
QY 256 ----- 256
Db 59806 CAGAAAGCTTGGAGCCAAAGTGGACAGAGGCGCTTGCCATTAAAGGCACTCGGCTCTGCCC 59747
QY 256 ----- 256
Db 59746 AGTGGAAAGGCGCTCCCTTCAGAAAGAAAGCTTCAGGCTGCCAGTGGGAAGGCGCTCC 59687
QY 256 ----- 256
Db 59686 TTGCAAGAAAGAAAGCTTCAGGCTGCCCTAGCTTCTTGCCATTCAGACCAACTCCTACAAAT 59627
QY 256 ----- 256
Db 59626 TACAAATTAGTGTCTTTTCCCTTCATTGGCTTTAAGGAAAGCTATTCAATTGCTGCTCG 59567
QY 256 ----- 256
Db 59566 CCAGTTCAGAGTAAAGTTGATTTGTAGGAGCCATAGCGGACGACGAAAGATCGCTGAT 59507
QY 256 ----- 256
Db 59506 TCCATAATGGGCGTAAGTACCGCGGCTTAACTTGTGTGTGAAGTCTGTACTTGTCACT 59447
QY 257 ----- 257
Db 59446 TCCCTCCAGCTCCCTGCTCGGATTTGAACGATGGAAGGCGAGTGGAGCTGTGTAAGAGG 59387
QY 274 AspAsnGluysGluArgHisLysLeuPheGlnGlyTyrGluThrGluGluArgGlu 293
Db 59386 GATPACGAGAAAGGAAAGCAAGCTGCTCCAGGCGTATGAACTGAAAGAGAGAGAGAA 59327
QY 294 ThrGluLeuSerGluLysIleLysLeuGluCysGlnProGluLeuSerGluThrSerGln 313
Db 59326 GCAGAGTGTCTGAGAAATTTAAATTGAGCGCGCAGCGGAGCTTTGCGAGACCTCCAG 59267
QY 314 ThleuProProLysPheSerCysGlyArgSerLysIleLysGlyValLys----- 330
Db 59266 GCTCTGCTTCCAAAGCCTTCTCTCATGTGCGGAGTGAAGGAGACCAAAAGGTGTGCT 59207
QY 330 ----- 330
Db 59206 GGTCACTTATTTCTGTTACTGAAAAAAGATGAGAGCTGCTAATGTGATGGAATATAGAG 59147
QY 330 ----- 330
Db 59146 CTTGAGGTTGATGATTGGATTTATGCAACATACCTGAATTGAGAAAGTCTGTATTC 59087
QY 330 ----- 330
Db 59086 TAGGGGAGTGTGCTTACCCCTGTATCAACAGCACTTAGAGTCTAAGCAAGATCT 59027
QY 330 ----- 330
Db 59026 GTTGTAGTTCAAGGTACGCTGCACTGAGGAACTGTGTTTTTACAAACAAAGAGCTTC 58967

QY 330 ----- 330
Db 58966 CAGGTTTCTGTGTGAGAAATGCTCTTGAAACATTTTCTTGTCACAGTACTAAGTACTC 58907
QY 330 ----- 330
Db 58906 GGGTGCAGCTACTCCCGAGAGGCGCTGCTGTATACTTTGAAGCTGACCTTGCTGC 58847
QY 330 ----- 330
Db 58846 CCATGGCTTCAACAGACAGATAGATATTTAGTACTGATCCTGATCTCTTGCCCTCAAA 58787
QY 330 ----- 330
Db 58786 TATATTCAATCAATTTATTTATTTGAATCAATTAATTTTGTGTTTGTGTTTGG 58727
QY 330 ----- 330
Db 58726 AGACCGGTTTCTCTGTAGACCAAGCTGGCTCGAACTCGAAATCTGCGCTGCTGCC 58667
QY 330 ----- 330
Db 58666 TCCCTAGTGTGGATTTAAAGGCGTGTGCCACCACTGCGCTCGAAATCAATTAAT 58607
QY 331 ----- 331
Db 58606 TTTATACATTTTTTTTCTACTTAATAGGAAACCCCATTTGGAATACAGAAAGAAAG 58547
QY 342 ThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysHis 361
Db 58546 ACTCCTGTTAAAGCTGGCTCCTGAATTTTCAAAAGTCAAAACAAACAACTCTTAAGCAC 58487
QY 362 SerProLysLysGluLysProCysGlySerLeuSerGluThrValCysLysArgGluLeu 381
Db 58486 TCTCCCATTAAGAGGAAACCTGTGGTCCATATTCAGAAACGTGTTGTAACGTGAATTG 58427
QY 382 ArgSerGlnGluThrProGluLysProArgSerSerValAspThrProProLysLeuSer 401
Db 58426 AGGAGCCAAAGAAACCCAGAAAGCCCGGTCTTCAGTGAATACCCACCAACACTCTCG 58367
QY 402 ThrProLysLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluLeu 421
Db 58366 ACTCCCAAAAGGAGCCAGACCCACCCAGAGAAAGACCTTCTCAAGTGAAGTAGAA 58307
QY 422 AspLeuProTyrLeuSerThrThrGluMetTyrLeuCysArgTyrHisGlnProProPro 441
Db 58306 GATTGGCGGTACTTCCACACAGAAATGTATTTGTGCTTGAGCACGAGCTCCGCCA 58247
QY 442 SerProLeuProLeuArgLysSerSerProLysLysGluLysValAla 458
Db 58246 TCACCGTTACATTACGGGAATCTCTCCAAAGAAAGAGAGACTGTAGCA 58196
RESULT 6
AC113356/c 183618 bp DNA linear HTG 18-SEP-2002
LOCUS Rattus norvegicus clone CH230-278A19, *** SEQUENCING IN PROGRESS
DEFINITION *** 7 unordered pieces.
AC113356
AC113356.2 GI:23101140
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae;
Rattus.
1 (bases 1 to 183618)
REFERENCE
AUTHORS Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas, V., Carter, K., Cavazos, I., Cesaar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carrillo, J., De Anda, C., Deedrick, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Diver, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durkin, C., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foerster, M., Foster, P., Fraser, C. M., Gabisel, A., Ganta, R., Garcia, A., Garner, J., Garza, M., Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowicz, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, Y., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzuewa, L., Louised, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mamoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Mielavjevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelmen, O., Okunodu, G., Olarunsaagun, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polndexter, A., Popovic, D., Prims, E., Pu, L.-L., Prazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojase, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shastman, S., Shen, H., Shetty, J., Shartsbeyn, A., Sison, I., Sitter, C. D., Sma's, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valdes, R., Vera, Y., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhuesern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 183618)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-Nov-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 183618)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (18-Sep-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 18, 2002 this sequence version replaced gi:22380622.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atl/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

-----Genome Center-----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GGOU
Center clone name: CH230-278A19

```
----- Summary Statistics -----  
Assembly program: Phrap; version 0.990329  
Consensus quality: 176229 bases at least Q40  
Consensus quality: 178023 bases at least Q30  
Consensus quality: 179050 bases at least Q20  
Estimated insert size: 200566; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
```

```
* NOTE: Estimated insert size may differ from sequence length  
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).  
* NOTE: This is a 'working draft' sequence. It currently  
* consists of 7 contigs. The true order of the pieces  
* is not known and their order in this sequence record is  
* arbitrary. Gaps between the contigs are represented as  
* runs of N, but the exact sizes of the gaps are unknown.  
* This record will be updated with the finished sequence  
* as soon as it is available and the accession number will  
* be preserved.
```

```
    1      35058: contig of 35058 bp in length  
    *      35059       35158: gap of unknown length  
    *      35159       139365: contig of 104207 bp in length  
    *      139366       139465: gap of unknown length  
    *      139466       140917: contig of 1452 bp in length  
    *      140918       141017: gap of unknown length  
    *      141018       142230: contig of 1213 bp in length  
    *      142231       142330: gap of unknown length  
    *      142331       145121: contig of 2791 bp in length  
    *      145122       145221: gap of unknown length  
    *      145222       148094: contig of 2872 bp in length  
    *      148095       148194: gap of unknown length  
    *      148195       183618: contig of 35424 bp in length.
```

```
Location/Qualifiers  
    1..183618  
     /organism="Rattus norvegicus"  
     /mol_type="genomic DNA"  
     /db_xref="taxon:10115"  
     /clone="CH230-278A19"  
     /note="wgs_contig"
```

```
ORIGIN  
misc_feature  
1..1263  
/note="wgs_contig"
```

```
Alignment Scores:  
Pred. No.:          Length:      183618  
Score:             Matches:        431  
Percent Similarity: 26.12%   Conservative:  8  
Best local Similarity: 25.64% Mismatches:   18  
Query Match:         Indels:       1226  
DB:                  Gaps:           5
```

```
US-10-054-935-2 (1-614) x AC131356 (1-183618)
```

```
Oy      15  AlAGlYGlYaMbProGluGiNaTGrLeAuAPrYtGUARAlaaLaaleuGIvGIyPro 34  
Db      67222 GCCGGCGGCACCCCGAGCAGCACCTGAACAAGAAGGGGCTTGCGCTTCGGCGGAGCCC 6716  
Oy      35  GUAsPglubProGIvAlaalagIuaLahIsphleuPrOAgrHIsArglyseulySGiu 54  
Db      67162 GAAGAGCAAGTCGGGGGGGCCGAACCACACTTCCTCCCCGGGATGTAAAGTCMAAGGAG 6710  
Oy      55  ProGIlYProBroleuALasErSerGIvGIySeRProlaPRoSeRProlaGIvCs 74  
Db      67102 CCCGGGGCCCCCGCTGCGCTTCACCAAGGGGGGAGAGCCCGCGCTTCACGCGGCTCC 6704  
Oy      75  ---GlvGYLYvAgEGLvArPGvlEuLeuLeuPrOalAGlVAlaalabProGIvGINginGu 93  
Db      67042 GGCGGCGGCAGAGGCGCGGGGTTGTACTCCCGCGCGGGGCGGCCCCCGGCGAGCAAGAA 6698  
Oy      94  GlUsEtTrIpJvGIySeRVAlProleuPrOCysPProProProalaThrlYsgInaIGly 113  
Db      66982 GAGAGGCTGGGGCGGTTCCGGGcCTTGccCTGncCGccccCGGCTAACAAAGCcGcg 6692  
Oy      114 IlEGlYGlYuBProAlaaLaagiAGlYCSeserPRoaGRproLystrYGlnaLa 133
```


QY 257 ----- 257
Db 64703 ATCTCAATACAAACAGAAATATATGTGATATATGTTGGATTATATTACTGAGGTGC 64644
QY 257 ----- 257
Db 64643 ATTATTACTGTTCCAGTGTAGTATGTTTAAAGTAAGTACTCCCTCAATACGACACATPAC 64584
QY 257 ----- 257
Db 64583 ATTTGAATGTGATCTGATCTGATCCCTAGGGAAGTCTGTGAGCTCTTGACTCTTCTTAC 64524
QY 257 ----- 257
Db 64523 TGTGCTGTTAAACATTCACAGTCTAGATTTTGGAGTGAACAGAGCAGATAGAA 64464
QY 257 ----- 257
Db 64463 TTTGGACACCGTGGGACACCTTAAAGAACACAGGGCTGAGAAATGATGCTCATTTGCTGC 64404
QY 257 ----- 257
Db 64403 TTAGAGTCTTGGCCAGACACGTTGGAGGCGCTGTGTTCTGTCCCTAGCACCATATCAA 64344
QY 257 ----- 257
Db 64343 CAGTGAAGAACATAGAGAGGTTCAATTGAGTGGCTCCAAAGACCTTGGAGCCACATGG 64284
QY 257 ----- 257
Db 64283 CTAGCTGACAGAGGCTTTCCTTCATTTAAAGGACCTGACCCCGCCAGTGGGGA 64224
QY 257 ----- 257
Db 64223 GGATCATCTTTACAGAAAGATGCTTGGATTGTCATCCCTTACCATCAGACTAA 64164
QY 257 ----- 257
Db 64163 CTCTATAATTCAATTGCTTTTGGCTTACTGSGTTTAAAGGAAGTATTCAGTGTGCC 64104
QY 257 ----- 257
Db 64103 TGCCAATTACAGAGAAAGTTGAATGTGATGGTCAACAGCAACATGAGAGGCCAAT 64044
QY 257 ----- 257
Db 64043 TCCTACATGGGGTTAAGTACTGTGCTCAATTTCTATGATATAGTCTGCATCTTGTGACT 63984
QY 258 ----- 258
Db 63983 TCCTTTAGCTCCTTGGCTCGGATTTGAAGTATGGAAGGCGGATGCGAGCTGTAAAGAG 63924
QY 274 AAspAsnGluysGluArgHisLysLeuPheGlnGlyTyrGluThrGluArgGluGlu 293
Db 63923 GATTAACAGAGAGAAAGCAAGCTGTTTCAGGGCTATGAACTGAAAGAGAGAGAA 63864
QY 294 ThrGluLeuSerGluLysLeuLeuGluCysGlnProGluLeuSerGluThrSerGln 313
Db 63863 ACGGAGTTGCTTGGAATAATTAACTGAGGCGCACCGGAGCTCTGGAACATCTCCAG 63804
QY 314 ThrLeuProLysProPheSerCysGlyYarSerGlyLysGlyHisLysArg----- 331
Db 63803 GCTTGCTCTTAAGCCTTTCTCATGTGGCCGAGTGGCAAGGACACAAAAGGTGTGCT 63744
QY 331 ----- 331
Db 63743 GGTGACTTATTTCTATGAACTGAAAAAGATAGAGAACTGTAGTCTGATGAAATATGCG 63684
QY 331 ----- 331
Db 63683 CTTTGATCTGTGATGGCTTTAGAGCAGCCATAGCTGAATTGAGAACTGTGTGTAC 63624
QY 331 ----- 331

Db 63623 TGGAGATGSGGCTTCCCTGTATTCACAGCACTAGTTCAAGGCCAGCCTGGCAGTAG 63564
QY 331 ----- 331
Db 63563 AGAACTGTTTTTACTTGTCTTGAGACATTTTTTTTTTTGTTACAGTGTCTTAAGTACAGANN 63504
QY 331 ----- 331
Db 63503 NNN 63444
QY 331 ----- 331
Db 63443 TTTTACATTCCTGTCAATGAAAAACAAGACCTTAGGTGGCCTAGCAGAGGGCT 63384
QY 331 ----- 331
Db 63383 CAAGACTCAGAGGCCGAGGCTCTGTGTGAAGAACTCCCTCTATTAATCTGGAAG 63324
QY 331 ----- 331
Db 63323 TTCCAAGCGTAACAGAACTTGAGTTGAGATTAAGACAGACTCTGCGAGACATCCAG 63264
QY 331 ----- 331
Db 63263 GCTCTGCTTCTAAGCCTTTCTCATGTGCGGAGTGGCAAGGACAAAGGTGTGCT 63204
QY 331 ----- 331
Db 63203 GGTGACTTATTTCTATGAACTGAAAAAGATAGAGAACTGTAAGTGTGATGAAATATGCG 63144
QY 331 ----- 331
Db 63143 CTTTGATCTGTGATTTGCTTTAGAGACACCATAGCTGAATTGAAAGTCTGTGTAC 63084
QY 331 ----- 331
Db 63083 TGGAGATGSGGCTTCCCTGTATTCACAGCACTAGTTCAAGGCCAGCCTGGCAGTAG 63024
QY 331 ----- 331
Db 63023 AGAAGCTGTTTTTACTTGTCTTGAGACATTTTTTTTTTTGTAGACGTTAAGTACAGAA 62964
QY 331 ----- 331
Db 62963 TGCAGATATTTTAAAGAAATGGAATAGGGGTTGGGATTTAGCTAGTGTAGAG 62904
QY 331 ----- 331
Db 62903 CGCTGCTGCAAGCTCAAGGCCCTGGTTGCGTCCGAGCTCCGAAAAAGAACAC 62844
QY 331 ----- 331
Db 62843 AAAAAAACAAGAAATAGTCTTAAGCATAGACGTTCTCATTTGACCAAGCCCTT 62784
QY 331 ----- 331
Db 62783 TCTGTAATCTTTAAGCACTGACACTTACTGTCCATGCGTTCCAAATGAACGTACATA 62724
QY 331 ----- 331
Db 62723 TCTTACATCTGCTACCTCTTGGCCCTCAAGTATATCAATCAAACTTATTGAA 62664
QY 332 ----- 332
Db 62663 ATTAATTAATTTTATTAACATTTCTTTCGTACTATATAGAAAAC-CAATTTGGAAT 62605
QY 338 ThrGluArgLysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLys 357
Db 62604 ACAGAAAGAAAGACTCTGTATTAAGCTGGCTCTGTAATTTTCAAAATTCAAAACAAA 62545
QY 358 ThrProLysHiserProLysGluGluProCysGlySerLeuSerGluThrValCys 377
|||||

```

Db      62544 ACTCTTACGACCTCTCCATTAAGAGAGACCCCTGTGCTTCATATCAGAAACTGTTGT 62485
Qy      378  TysArgGlnLeuArgSerGlnGlnJuthrProGlnLysProArgSerValaIapThrPro 397
Db      62484 AAACGTGAATTGAGGACCAAGAAACCCCAAGAAAGCCCGGTCTTCAGTGATACCCCA 62425
Qy      398  ProArgLysSerThrProGlnLysGlyProSerThrHisProLysGlnLysAlaPheSer 417
Db      62424 CCAGAGCTCTCGACTCCCAAGAGGAGCCAGACCCAGACCCCAAGAGAGGAGCTTCTCA 62365
Qy      418  SerGlnIleGlnLysLeuPProTyrLeuSerThrThyGluMetYrLeuCyGTAAGTTPHIS 437
Db      62364 AGTGAGATGAGAGATTTCCTGCTTCCTTCACACAGAAATGATTTGTCTTGGCAC 62305
Qy      438  GlnProProSerProLeuProLeuArgGlnLysSerProLysGlnGlnJuthrVal 457
Db      62304 CAGGCTCCCCCATCACCGTTACCATTAACGGAAATCCTCTCCAAAGAGAGGAGACTGTA 62245
Qy      458  AlaArgCysLeuMetProSerSerValAlaGlyLuthrServalLeuAlaValProSer 477
Db      62244 GCA-----AGTAAGGCTAGCGAAGACTTCTCTT---ATACCTAG 62206
Qy      478  TTP 478
Db      62205 TGG 62203

RESULT 7
LOCUS   AC068669 194780 bp DNA linear PRI 17-JAN-2003
DEFINITION Homo sapiens chromosome 17, clone RP11-749116, complete sequence.
ACCESSION AC068669
VERSION   AC068669.21 GI:27777612
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eureleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS  Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
          Bogenbas,A., Baldwin,J., Barna,N., Baetsen,V., Beda,F.,
          Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G.,
          Campioiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
          Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
          Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
          Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
          Grand-Pierre,N., Grant,G., Hages,B., Heaford,A., Horton,L.,
          Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatae,A.,
          Levine,R., Licoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
          Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
          McElrath,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
          Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
          O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
          Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
          Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
          Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
          Testfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,
          Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
          Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE     Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
JOURNAL   Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
          Barna,N., Baetsen,V., Bloom,T., Boguslavsky,L., Bouckhalter,B.,
          Camarata,J., Chang,J., Chazaro,B., Choquel,Y., Collymore,A.,
          Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
          Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
          Hages,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatae,A., Kells,C., Landers,T., Levine,R.,
          Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
          Mathews,C., McCarthy,M., Meldrum,J., Menus,L., Mihova,T.,
          Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
          Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
          Peterson,K., Phunhthang,P., Pierre,N., Raymond,C., Retta,R.,
          Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupbach,R.,
          Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
          Stojanovic,N., Talamas,J., Testfaye,S., Theodore,J., Topham,K.,
          Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
          Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE     Submitted (17-JAN-2003) Whitehead Institute/MIT Center for Genome
JOURNAL   Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
          Barna,N., Baetsen,V., Bloom,T., Boguslavsky,L., Bouckhalter,B.,
          Camarata,J., Chang,J., Chazaro,B., Choquel,Y., Collymore,A.,
          Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
          Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
          Hages,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatae,A., Kells,C., Landers,T., Levine,R.,
          Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
          Mathews,C., McCarthy,M., Meldrum,J., Menus,L., Mihova,T.,
          Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
          Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
          Peterson,K., Phunhthang,P., Pierre,N., Raymond,C., Retta,R.,
          Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupbach,R.,
          Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
          Stojanovic,N., Talamas,J., Testfaye,S., Theodore,J., Topham,K.,
          Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
          Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

```

```

Hages,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatae,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Mathews,C., McCarthy,M., Meldrum,J., Menus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunhthang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Testfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE     Submitted (17-JAN-2003) Whitehead Institute/MIT Center for Genome
JOURNAL   Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
          Barna,N., Baetsen,V., Bloom,T., Boguslavsky,L., Bouckhalter,B.,
          Camarata,J., Chang,J., Chazaro,B., Choquel,Y., Collymore,A.,
          Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
          Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
          Hages,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatae,A., Kells,C., Landers,T., Levine,R.,
          Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
          Mathews,C., McCarthy,M., Meldrum,J., Menus,L., Mihova,T.,
          Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
          Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
          Peterson,K., Phunhthang,P., Pierre,N., Raymond,C., Retta,R.,
          Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupbach,R.,
          Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
          Stojanovic,N., Talamas,J., Testfaye,S., Theodore,J., Topham,K.,
          Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
          Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE     Submitted (17-JAN-2003) Whitehead Institute/MIT Center for Genome
JOURNAL   Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
          Barna,N., Baetsen,V., Bloom,T., Boguslavsky,L., Bouckhalter,B.,
          Camarata,J., Chang,J., Chazaro,B., Choquel,Y., Collymore,A.,
          Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
          Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
          Hages,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatae,A., Kells,C., Landers,T., Levine,R.,
          Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
          Mathews,C., McCarthy,M., Meldrum,J., Menus,L., Mihova,T.,
          Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
          Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
          Peterson,K., Phunhthang,P., Pierre,N., Raymond,C., Retta,R.,
          Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupbach,R.,
          Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
          Stojanovic,N., Talamas,J., Testfaye,S., Theodore,J., Topham,K.,
          Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
          Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

```

```

FEATURES
source
1..194780
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-749116"
/clone_lib="RP11-749116 Human Male BAC"
2..285
/rpt_family="L1MC3"
302..324
/rpt_family="(TTG)n"
529..637
/rpt_family="FLAM_A"
548..553
/note="<30 qual SINGL region"
complement(711..1007)
/rpt_family="AluSx"
complement(1038..1179)
/rpt_family="MER115"
2169..2473
/rpt_family="AluSp"

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10045
Center clone name: 749_I_16
-----
Location/Qualifiers
1..194780
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-749116"
/clone_lib="RP11-749116 Human Male BAC"
2..285
/rpt_family="L1MC3"
302..324
/rpt_family="(TTG)n"
529..637
/rpt_family="FLAM_A"
548..553
/note="<30 qual SINGL region"
complement(711..1007)
/rpt_family="AluSx"
complement(1038..1179)
/rpt_family="MER115"
2169..2473
/rpt_family="AluSp"

```


Db	68364	TTATCTGCATTCGGGCGCCAGAGCAGGATGTTGCGCATGCTCGGGGAAAGGGGCTTAG	68423
QY	256	-----	256
Db	68424	GAGGTTAAGCACCCCGGGTTAGGGGTAAGAAGGTTGGCCACCGCAAGAGATCT	68483
QY	256	-----	256
Db	68484	TAGCGCTGGGTCTCTGGAGCCAGGCTCACAGACAGTTGGGTTGAGGAAAGGTTAA	68543
QY	256	-----	256
Db	68544	AGGCAACCTCTCAGGGGGAGGGGGAGGGGTTTCAGGTGCMAAGGTTAATTTAAATG	68603
QY	256	-----	256
Db	68604	ACAGCCCAAGCGTGTGGAAAGCGTTTAGTGAAAAAGACGTGATCTGTGAGGGTTA	68663
QY	256	-----	256
Db	68664	AAAATGAAGGGTGGGTGATGTTGTCTCTATGAGGGTTAAATTAAGCTA	68723
QY	256	-----	256
Db	68724	TTGGCTTAGAAATCAGCGGTGAGAGAGGGGCTGACAAAGACGTGTCTTCAACAC	68783
QY	256	-----	256
Db	68784	TGCTGAGCATCTTTAGATAGAGGGCAGTAGGCTTTTAGTGAGCAGGGGTAGAGGT	68843
QY	256	-----	256
Db	68844	GATTTAGGTGAATCCTATTTTTACTTTTCTTAAATGATTTAAGAAATCATCGCTTGG	68903
QY	256	-----	256
Db	68904	ATTTCAAGACGTTTAAATAGAAACGAGTTCTTAATACATGACCACTGACCAAGGTCC	68963
QY	256	-----	256
Db	68964	TTTTATAAGCATGGTGTGTGAACTGGAGATTTTAGGGCATGAGGAAATGAACTTA	69023
QY	256	-----	256
Db	69024	TTAATGATTAATAATAGACCTCCCATCTTATTTTTGAATCTTAGGAGTAGAAAAATC	69083
QY	256	-----	256
Db	69084	GTTATTATTCTTTACTTCTGAGAGAGCTTGAGACAGACAGGCTGTGTTTTACTGACGA	69143
QY	256	-----	256
Db	69144	GTAAGTAAGGCATCTCGGTTAAACAATATTCCTTCAATATGAGGAAAGCTGGGTTG	69203
QY	256	-----	256
Db	69204	GGGGATTAATCTGTGTAGCATTAACAGCTGCTAAAAAATPAACATGTTTCTACAA	69263
QY	256	-----	256
Db	69264	TTGTGTGTGAGGCTGAGAGCATCTCTCTCCCTCTCCCTCCCTCTCTCTCCCTCC	69323
QY	256	-----	256
Db	69324	CTGTCTCTTCCTTCTCTCTCTCTCTCTCTCTCCCTCCCTCCCTCCCTCATCTCTC	69383
QY	256	-----	256
Db	69384	CCTCTCTCTCTTCTCTCTGCGGTGCTGTCGCTAAGACTGAGCAAGGAGCAAGCG	69443
QY	256	-----	256

Db	69444	CTATTCAATTTTGGTTGGCTTTCCAAATGAAATCTACGCGGCGTGTTTTATGAGAAATG	69503
QY	256	-----	256
Db	69504	CATTCTGCAAGTGAATTGGAGCTGGGGCAGGTCTCATTTAAATATCTCATGGTGTGA	69563
QY	256	-----	256
Db	69564	TTTGTGGAAGTTACTCAACATGAGGTGGCACTGCTCAGACTGTACTGACGTAGGGGT	69623
QY	256	-----	256
Db	69624	GTGATGTAACCGTAATTACAAACATCATATCCTTGTATGGGGGAGTGGGAAGAAGG	69683
QY	256	-----	256
Db	69684	ACCAAGTAGAGAAAAAATCAAGAAATGTCACCACAATGTGTGTTAAGGCTGTG	69743
QY	256	-----	256
Db	69744	AATCAAAAATATATCAAGAGATGAATGAGAGTATGGCTGATAGGGGGTTCTCTGCTG	69803
QY	256	-----	256
Db	69804	ATTGTGAGGTTTCTGCTCACTTCCTGAGTAATAGCCCCCTTGCAGGCTTTCCTAGTCA	69863
QY	256	-----	256
Db	69864	GAAAAAGTGGGGTAGTTAAGATAGAGTGTCAGTCTGAACCTTGTACTCACTGCTCTG	69923
QY	256	-----	256
Db	69924	TTGTGTATGCAATCTTAATATGTGCTTAAGCTGTTCATCTTAATCCAGTTGTGTTGG	69983
QY	256	-----	256
Db	69984	CCTGTCTGAATCTATGCAATGATGTAAATGAAGCAGCGTTAAACATCACTGTCTAAC	70043
QY	256	-----	256
Db	70044	CCAGTAATAATCTTGTATATAGATATATAGACACACACCTTTTCTCAAAACTAACT	70103
QY	256	-----	256
Db	70104	AACACAAAGTTTCAATACTAATAGAGTTCATTTCTGTTAGTTTATAGTCTTCT	70163
QY	256	-----	256
Db	70164	TTTAACATAAACTAAGAACCTTGATTTTAATAATTTAGATGAATAGTAAATTT	70223
QY	256	-----	256
Db	70224	TTTGTGTACTCAAAATTCAAAGAAATGTGTTATGTCATGATGTGGAATATAGGTA	70283
QY	256	-----	256
Db	70284	CTGATTTCTATTTTCCAATGAATAATGTTCATTTGTTTAAAGTACCCCTCCACACATA	70343
QY	256	-----	256
Db	70344	CCCTTTCAAAATACACATGCCATTCCTGTGTGCACTGTATGAGCTTTAATTTCTTCTCA	70403
QY	256	-----	256
Db	70404	ATGTACTGTTTGCACCTTGTGTTGTGTATAGAGCTTTGAGATAGAAGTAGAGAGATAG	70463
QY	256	-----	256
Db	70464	AATTTGACAAATTCAGGGCTCTTAAAGATCATATGAGCTGGCGGGTGTCTACGCG	70523
QY	256	-----	256
Db	70524	CTGTATATCCAGCACTTTGGAGGCGCTTAGCGAGGTGACACAGAGGTCAAGATAGAGA	70583

QY	256	-----	256	-----	330	-----	330	-----
Db	70584	CCATCTGGCCACATGATGTAACCTGTCTCTACTAAAAATACAAAAATTACCCGGTGT	70643		71664	AGTACTCGGAGGCTGAGGACAGAGAATCACTTGAACCCGGAGGCGAGTTGACATG	71723	
QY	256	-----	256	-----	330	-----	330	-----
Db	70644	GGTGGCGGGCGCTGAATAACCACTACTCGGAGGCTGAGGACGAGGCGGAGGTTGC	70703		71724	AGCCGGATTGCTCCACTGCATCTCAGCCTGGCAACAGAGGAGATCCGTCAAAAAAA	71783	
QY	256	-----	256	-----	330	-----	330	-----
Db	70704	AGTGAGCATTTGCAATCAGCCTGGCAACAGCAAACTCCGTCCTCAAAAACAAAC	70763		71784	AAAAAAAAAGAGTCTTCTATGCTGAGAACTATCTTGAACATTCTTATGACAGCTGTA	71843	
QY	256	-----	256	-----	330	-----	330	-----
Db	70764	AAAAACAACAAACAAACATGTATACATTTATTTCTGAAACCTCCGAAACCACTTGA	70823		71844	AAGTCTTTAGAGTATTTTGTGACAGTAATATTTGAATGACCTTACATCATTA	71903	
QY	256	-----	256	-----	330	-----	330	-----
Db	70824	GCCACACAGGTAGCATGACAGACCATGCAATTAAGGCATGTTACGCCCTTCCA	70883		71904	AGAGATGGCAATTAAGTATGACAGTGAAGTTCTTAAAACTATAGCAGTTCACTGAA	71963	
QY	256	-----	256	-----	330	-----	330	-----
Db	70884	GTGAGAGGATCTCCCTACAGAGAAATGAACATAATCCCATGCTCTTCCACTCT	70943		71964	TTAGAAAGTGTAAATTCCTTCATTGTGTGCTGATTTGTAGACATAGAAAAATTAATA	72023	
QY	256	-----	256	-----	330	-----	330	-----
Db	70944	AGGTAGCCCTTAATTTAACATATACCTCTCTCTCTATTTTGTTAAGAAAACT	71003		72024	CAGGCGAGCGGTGTGCTCATGCTGTAAATCCCTTGAGAGGGCAAGTGGGAGAG	72083	
QY	256	-----	256	-----	330	-----	330	-----
Db	71004	GTTCAGATTAATGCTGCTCTATTTAAAAAGTTGATGTGTGTCATAGAAAGGCA	71063		72084	TATGAAAGTATGAAAGTGGGAGAGGAGATCGCTTGGCCGAGATTGAGACCG	72143	
QY	256	-----	256	-----	330	-----	330	-----
Db	71064	GAAAGCTATATGATCTCTTAAGTGGGCTAAGTATCTGTGTGTTAATTTTATGTGT	71123		72144	CTGGGCAATATGATGAGACCTCATTTTACAAAAATCCAAAAATTACCGGCTTGA	72203	
QY	257	-----	257	-----	330	-----	330	-----
Db	71124	AAGCTGCATTTTGTCTACTCTCTCTTTAGCTCTTGTCTGGAATGGAACGATGAAAG	71183		72204	GGTATGACCTGTATGCTCCAGCTACTCAGAGGTTGAGGCTACAGTAGCTTACTGA	72263	
QY	267	ArgMetGlnLeuValIysIysASPasnGluIysGluIuArgHisIysLeuPheGlnGlyTyr	286		72264	CAACTGCATCCAGCCTGGGTGACGAGTATACCTGTCTCAATGAAAAA	72323	
Db	71184	CGGATGAGCTGTAAAGAGATTAACGAGAAAGGACAGCTGTTCAGGGCTAT	71243		330	-----	330	-----
QY	287	GluThrGluGluArgGluIuThrGluLeuSerGluIysIleIysLeuGluIuGlnPro	306		72324	AAAAAGTAAATGAACTGTCTCTTGACTCAAAAAATTAGCCAACTGTAGGATAG	72383	
Db	71244	GAAACTAAGAGAGAGAGAAACAGACTATCTGAGAAATTAACCTGAGATGCCAGCG	71303		330	-----	330	-----
QY	307	GluLeuSerGluThrSerGlnThrLeuProIysPheSerCysGlyArgSerGly	326		72384	CATACATTTTGAACACTCAGATTTCAATGCAATGCAATGAAATAGTATGATACACACCTGA	72443	
Db	71304	GAGCTTCCGAGACATCCAGACTCTGCTCCAAAGCCTTCTCATGTGGCGGAGTGA	71363		330	-----	330	-----
QY	327	IysGlyHisIys	330		72444	TATTAAGCTTTTGTGTTGTGATAGAGGTCTCTCTGTACCCAGGCTGGGG	72503	
Db	71364	AAGGACATATAAAGGTGCTGATACCTTTGATGAGGACCTTGTTAACAGTAT	71423		330	-----	330	-----
QY	330	-----	330	-----	72504	TGCAAGTGGCCAAATGGCTCATCTTCACTCTTGACCTCAAGTATCCTCC	72563	
Db	71424	GAGAGTAGAGATTTGATTTGATTAATGTTTATAGCAATCAATCTGCTTTAGAG	71483		330	-----	330	-----
QY	330	-----	330	-----	72564	TACTTACGCTCCCATGTAGCTAGACCATAGGACATCACCATGCTTGGCTACTTTT	72623	
Db	71484	TCTTCTATGCGAGGCGCGGCTGTGCTCACACTGTATATCCAGCACTTTGGAGG	71543		330	-----	330	-----
QY	330	-----	330	-----	72624	TAGTTTGTAGAGACAGCTCTCATCTTGTTCAGGCTGATCTCAAACTCTGGGCT	72683	
Db	71544	CCAAAGCGGATGATCATGAGGTCAAGAGATAGAACATCTGGCCAAATTTGAAC	71603		330	-----	330	-----
QY	330	-----	330	-----	72684	CAAGTATCTCCGAAAGTGTGGATTAACAGACATTAAGTTGTGAATTCAG	72743	
Db	71604	CCCGTCTCTATAAAAAATATAAAAAATTAGCCGGGCGTATGCGGCGCCTGTATGCC	71663		330	-----	330	-----

Db 72744 GGATATTAAGTATTTCTGGTTAGAAATTTGAAACTTTAGTAGAAATCACTTTGAA 72803
QY 330 ----- 330
Db 72804 AAGAAGTATAGCGGGGTGCGTGTCTCACGCTGTATATCCAGCACTTTGGAGGCCGA 72863
QY 330 ----- 330
Db 72864 GCGGGGTGATACGAGGTGAGAGATCGAGACATCTGGCTTAACACGGTGAACCTTG 72923
QY 330 ----- 330
Db 72924 TCTCTAATAAATAAGAAAAATTAAGCCGGGTGTGTGATACACTGTAGTCCAGCTAC 72983
QY 330 ----- 330
Db 72984 TTGGAGGCTGAGGAGGAAATTTGCTTAACCCAGAGGTGAGGTTGCAGTAAGACA 73043
QY 330 ----- 330
Db 73044 GATCGACCACTGCACTCTCTGGGTGAAGAGTGAAGATCCAAAAATTAATAAT 73103
QY 330 ----- 330
Db 73104 TTAGTTGGTAAAGAGAGGAGATGACTTTCTTGGGATGTGTGGGGAAGAGGAA 73163
QY 330 ----- 330
Db 73164 GAAGAAAGAGCTAGAGAAATGCCAGACAGATAGAAAAAAGAAATTAAGTTACAAT 73223
QY 330 ----- 330
Db 73224 AATGTGTAAGGGTTGGGGAAAAAGAAAGAGACTGATCTGTGATTACAGAAAGAG 73283
QY 330 ----- 330
Db 73284 TGCACTAATGTGGTTGTAGGTACATATGTGGCTAAAGAGCCTTGAATATTTCTT 73343
QY 330 ----- 330
Db 73344 TTTTTTTTTTTTTTTTTTCTGAGACAGAGTCTCACTCTGTGCGCCAGGCTGAGTGCAT 73403
QY 330 ----- 330
Db 73404 GGGGTATCTAGCTCACTGCAACTCTGCTCCAGGGTCAAGCGATTTCTCTGCTCA 73463
QY 330 ----- 330
Db 73464 GCCTCTCAAGTACGTAGATTACAGCATGTGCCACCAAGCTGACTAATTTTGTATTT 73523
QY 330 ----- 330
Db 73524 TTTTTTTTTAGTAGAGAGCGGGTTTCAACATATTGACAGAGTGTCTGAACTCTGAC 73583
QY 330 ----- 330
Db 73584 CTCAGTGAATCCGCTGATTTGGCTTCAAAAGTGTGGATTACAGCGTGAAGCCACTG 73643
QY 330 ----- 330
Db 73644 CGCTGTGTGACCTTGAAATTTCTTAAGCATGAAATTCCTTGTCAACATTTACTGGCC 73703
QY 330 ----- 330
Db 73704 AGGTGAGTGTGCTCATGCGGGTATATCCAGACATTTGGAGGCTAGTGGTGTGATCAC 73763
QY 330 ----- 330
Db 73764 CTGAGGTCAAGTATTTGAGACAGTCTGACCAATATGTGAAAACCTGCTTATTAATA 73823
QY 330 ----- 330

Db 73824 ATACAAAAAAATTAACAGGCAATGTTGGTGTATGCTGTAAATCCAGGCTACTCAGAGG 73883
QY 330 ----- 330
Db 73884 CTGAGATAGAGAAATTTCTTGAACCCGGAGGAGAGATTTGAGTGAAGCCGATATGC 73943
QY 330 ----- 330
Db 73944 CACTGACTTCAGCTTTGGGTGACAGACAAAGTCTGTCTCAAAAAAATAAAACAAA 74003
QY 330 ----- 330
Db 74004 ACAAAAAATTTACTATGGGCTATTTTATATCTGTAAGCCTGGCTGTCTGCTCAT 74063
QY 330 ----- 330
Db 74064 GACTGCTAAATGACACACAGATTTAGGTACTGTAATCTTGTCAATTTTGGGCAAAA 74123
QY 330 ----- 330
Db 74124 AACCACTTATTTCTTGTCTCCAGAAATGTGTCAAAATTAAGATTATTCAGTGTCAATA 74183
QY 331 ----- ArgLysSerProPheGlySerThrGluArg 340
Db 74184 AATTTATATATGTTTCTTCCCTATCTATATAGAAATCCCAATTTGGAGTACAGAAAGA 74243
QY 341 LysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysProLys 360
Db 74244 AAGACTCTGTAAAGAGCTGGCTCTGAAATTTTCAAAAGTCAAAACAAAAATCTCTAAG 74303
QY 361 HisSerProLysLeuGluProCysGlySerLeuSerGluThrValCysLysArgLys 380
Db 74304 CACTCTCATTTAAAGAGAACCTGTGTCTTATCTGAACCTGTGTATAACGTGAA 74363
QY 381 LeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProArgLys 400
Db 74364 TTGAGAGCCAAAGAACCCAGAGAAAGCCCGGTCTTCACTGAGACACCCAGCAAGCTC 74423
QY 401 SerThrProGlnLysGlyProSerThrHisProLysGluValAlaPheSerSerGlu 420
Db 74424 TCCACTCCCAAAAGAGAGACCCAGCACCATCCCAAGAGAAACCTTCTCAATGAGATA 74483
QY 421 GluAspLeuProLysLeuSerThrThrGluMetTyrLeuCysArgThrPheGlnProPro 440
Db 74484 GAAGATTGGCGTACCTTTCCACACAGAAATATTTGTGTGGACACAGCTCC 74543
QY 441 ProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrValAlaArgCys 460
Db 74544 CCATCAACGTTACATTAACGGAAATCTCTCCAAAGAGAGAGAGACTGAGCA----- 74597
QY 461 LeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrp 478
Db 74598 -----AGTAAGGATAGAGAACACTGTCTT---ATACCCTAGTGG 74636
RESULT 8
CQ724830 655 bp DNA linear PAT 03-FEB-2004
LOCUS CQ724830
DEFINITION Sequence 10764 from Patent WO02068579.
ACCESSION CQ724830
VERSION CQ724830.1 GI:42285687
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLES Kils, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 10764 06-SEP-2002;
FEATURES Location/Qualifiers

Source

1. .655
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.:	6,7e-26	Length:	655
Score:	1116.00	Matches:	217
Percent Similarity:	76.68%	Conservative:	0
Best Local Similarity:	76.68%	Mismatches:	0
Query Match:	34.47%	Indels:	6
		Gaps:	1

US-10-054-935-2 (1-614) x CQ24830 (1-655)

```
QY 332 LysSerProhegIysSerThrGluArgLysThrProValIysLysLeuAlaProGluPhe 351
DB 2 AAATCCCATTTGGAAAGTACGAAAGAAAGACTCTCTTAAAGAGCTGGCTCTGAAATT 61
QY 352 SerLysValIysThrLysThrProLysHisSerProIleLysGluGluProCysGlySer 371
DB 62 TCABAAATCAAAACAAAACCTCTTACGACCTCTCTTAAAGAGAACTGTGTTCC 121
QY 372 LeuSerGluThrValCysLysArgGluLeuArgSerGlnGluThrProGluLysProArg 391
DB 122 TTATCTGAAACTGTTTGTAAACGTGAATTGAGAGCCAGAAACCCCAAGAAACCCCGG 181
QY 392 SerSerValAspThrProProLysLeuSerThrProGluLysGlyProSerThrHisPro 411
DB 182 TCTTCAAGTGAACACCCACCAAGACTCTCCATCCCAAAAGGACCCAGACCCATCC 241
QY 412 LysGluLysAlaPheSerSerGluIleGluAspLeuProTyrLeuSerThrThrGluMet 431
DB 242 AAGGAGAAAGCTTCTTAAGTGAATGAAAGATTGGCGTACCTTTCACCAAGAAATG 301
QY 432 TyrLeuCysArgTprHisGlnProProProSerProLeuProLeuArgGluSerSerPro 451
DB 302 TATTGTGCTGGTGGACCAAGCCCTCCCATCACGTTTACATTACGGGAATCTCTCCA 361
QY 452 LysLysGluGluThrValAlaArgCysLeuMetProSerSerValAlaGlyGluThrSer 471
DB 362 AAGAGAGAGAGACTGTAGCAAGGTGTCTGATGCATCAAGTGTGACGAGAAACTTCA 421
QY 472 ValLeuAlaValProSerTprArgAspHisSerValGluProLeuArgAspProAsnPro 491
DB 422 GTCTGGCTGTTCTCTTGGAGGACCACTCAGTAGGCTCTTAAGGAGCCCAATCT 481
QY 492 SerAspLeuGluGluAsnLeuAspAspSerValPheSerLysArgHisAlaLysLeuGlu 511
DB 482 TCAGACTTTTGGAGAACTGTAGTGAAGTGTCTTTCGAAGCGGCATGCAAACTGGAG 541
QY 512 LeuAspGluLysArgArgLysArgTprAspIleGlnArgIleArgGluGlnArgIleLeu 531
DB 542 CTGATAGAGAGAGAGAGAAAGG----- 565
QY 532 GlnArgLeuGlnLeuArgMetTyrLysLysLysGlyIleGlnGluSerGluProGluVal 551
DB 565 ----- 565
QY 552 ThrSerPhePheProGluProAspAspValGluSerLeuMetIleThrProPheLeuPro 571
DB 565 ----- 565
QY 572 ValValAlaPheGlyArgProLeuProLysLeuThrProGluAsnPheGluLeuProTyr 591
DB 566 -----AAATTTGAGCTCCTCGG 583
QY 592 LeuAspGluArgSerArgCysArgLeuGluIleGlnLysLysGluThrProHisArgThr 611
DB 584 TTGGATAGAGCTAGCCCATGCAATTTGAGATCCAGAGAGCAAAACCTCACCGAGCG 643
QY 612 CysArgLys 614
```

DB 644 TGTAGGAAA 652

RESULT 9

AX401049

615 bp

DNA

linear

PAT 06-JUN-2002

DEFINITION

Sequence 725 from Patent WO0210453.

ACCESSION

AX401049

GI:21337229

VERSION

AX401049.1

GI:21337229

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Rattus norvegicus

REFERENCE

Mendrick, D., Porter, M.W., Johnson, K.R., Castle, A.L. and

AUTHORS

Elashoff, M.R.

TITLE

Molecular toxicology modeling

JOURNAL

Patent: WO 0210453-A 725 07-FEB-2002;

FEATURES

Gene Logic, Inc. (US)

SOURCE

Location/Qualifiers

ORIGIN

1. .615

Alignment Scores:

Pred. No.:	4.03e-24	Length:	615
Score:	1055.50	Matches:	197
Percent Similarity:	97.55%	Conservative:	2
Best Local Similarity:	96.57%	Mismatches:	4
Query Match:	32.60%	Indels:	1
		Gaps:	1

```
QY 13 AlaProAlaGlyGlyAsnProGluGlnArgLeuAspTyrGluArgAlaAlaLeuGly 32
DB 3 GCCCTGCGCGGCGGCAACCCCGAGCAGACTGACATGACGCGGCTGCGCTGGC 62
QY 33 GlyProGluAspGluProGluValAlaAlaGluAlaHisPheLeuProArgHisArgLysLeu 52
DB 63 GGGGCCAGAGCAGTCCGGGGCGCGCGAGCCCACTTCTCCCGGCACTGTAAGCTC 122
QY 53 LysGluProGlyProProLeuAlaSerSerGlnGlyLysSerProAlaProSerProAla 72
DB 123 AAGAGCCCGGCGCGCGCGCTGCTTACCAAGGCGGAGCCCGGCGCTTCCAGCT 182
QY 73 GlyCys---GlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGlyGln 91
DB 183 GGTGCGCGCGCGGCAAGGCGCGGCGGTTGTATCTCCGCGCGGCGCGCGCGCGGCGAG 242
QY 92 GlnGluGluSerTprGlyGlySerValProLeuProCysProProProAlaThrLysGln 111
DB 243 CAGGAGAGAGCTGGGGCGGCTTCCGCTTCCCTGCTGCGCGCGCGCGCGCTACCAACA 302
QY 112 AlaGlyIleGlyGlyGluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTyr 131
DB 303 GCGGCGATCGCGCGGAGCCAGTGCAGCGCGCGCTGTCGACCCCGGCGCAAGTAT 362
QY 132 GlnAlaValLeuProIleGlnThrGlySerLeuValAlaAlaLysGluProThrPro 151
DB 363 CAGCGGTGTCTGCCATTCAGACGGGCTCTCTGTGCGCGCGCGCAAGAGCTTACCC 422
QY 152 TrpAlaGlyAspLysGlyLysAlaAlaSerProAlaAlaThrAlaSerAspProAlaGly 171
DB 423 TGGGCTGGGAGCAAGGTTGGGCGCGCTCCCGGCTCCACCGCTCGAGCCGCGGGA 482
QY 172 ProProProLeuProLeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThr 191
```

Db 483 CCCCCACCACTACTGCGCCGGGCGCCACCCCTGCGCCGCCACCGCACTGCGGGGACC 542

Qy 192 leuAlaAlaSerGluGlyArgTyrPlySerMetArgLysSerProLeuGlyGlyGly 211

Db 543 CTGGCGGCCAGTGAAGGCGCAGATGGAAGATATAGAAAGAGCCCTCGGGGGTGGCGGC 602

Qy 212 GlySerGlyAla 215

Db 603 NGCTCGGGAGACC 614

RESULT 10

AR379777 1290 bp DNA linear PAT 18-DEC-2003

LOCUS AR379777 Sequence 322 from patent US 6607879.

DEFINITION AR379777

ACCESSION AR379777

VERSION AR379777.1 GI:40087411

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1290)

AUTHORS Cocks,B.G., Stuart,S.G. and Seilhamer,J.J.

TITLE Compositions for the detection of blood cell and immunological response gene expression

JOURNAL Patent: US 6607879-A 322 19-AUG-2003;

FEATURES

source Location/Qualifiers

1..1290

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 8.65e-24 Length: 1290

Score: 1053.50 Matches: 207

Percent Similarity: 93.69% Conservative: 1

Best Local Similarity: 93.24% Mismatches: 6

Query Match: 32.54% Indels: 8

DB: Gaps: 2

US-10-054-935-2 (1-614) x AR379777 (1-1290)

Qy 258 leuAlaArgIleGluArgMetGluArgArgMetGluLeuValLysLysAspAsnGluLys 277

Db 3 CTTCCTCGAGTTAAAGTATGAAAGCGGAGCGAGTCACTGTTAAAGAAATACGAGAAAA 62

Qy 278 GluArgHisLysLeuPheGlnGlyTyrGluThrGluGluArgGluGluThrGluLeuSer 297

Db 63 GAAAGCACACAGCTGTTTCAGGGCTATGAAGCTGAAGAGAGAGAGAAACAGAGCTATCT 122

Qy 298 GluLysIleLysLeuGluCysGlnProGluLeuSerGluThrSerGlnThrLeuPro 317

Db 123 GAGAAATTAATACTGAGTGCCAGCCGCGAGCTTCCAGACATCCCGACTGCGCTCCC 182

Qy 318 LysProPheSerCysGlyArgSer-GlyLysGlyHisLysArgLysSerProPheGlySe 337

Db 183 AAGCCCTTCACTGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 242

Qy 337 rThrGluArgLysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLy 357

Db 243 TACAGAAAGAAAGAACTCTGTTAAAAAGCTGGCTCCGGAATTTTCAAAAATCAAAAACAA 302

Qy 357 rThrProLysHisSerProIleLysGluGluProCysGlySerLeuSerGluThrValCy 377

Db 303 AACTCTAAGCACTCTCTTAATAAGAGAAACCTGTGGTTCCTTATCTGAACCTGTTTG 362

Qy 377 sLysArgGluLeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrPr 397

Db 363 TAAACGTGAATTAGGAGCCAGAAAGCCAGAAAGCCCGGCTTCAGTGGACACCCC 422

Qy 397 oProArgLysSerThrProGluLysGlyProSerThrHisProLysGluLysAlaPheSe 417

Db 423 ACCAAGACTCTCACTCCCAAAAGGAGCCAGACACCATCCCAAGAGAAAGCTTCTC 482

Qy 417 rSerGluIleGluAspLeuProThrLeuSerThrThrGluMetTyrLeuCysArgTrpHi 437

Db 483 AAGTGAGATGAAGATTTGCGCACTTTCACACAGAAATATATTGTGTGGCA 542

Qy 437 sGlnProProPheSerProLeuProLeuArgLysSerSerProLysGlyGluThrVa 457

Db 543 CCAAGCTCCCGCATCAACGTTACATTACGGGAATCTCTCCAAAGAGAGAGAGACTGT 602

Qy 457 lAlaArgCysLeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSe 477

Db 603 AGCA-----AGTAAGCATAGAGAACACTTCTCTT---ATACCTTA 641

Qy 477 rTTP 478

Db 642 GTGG 645

RESULT 11

BC039449 1269 bp mRNA linear PRI 21-OCT-2003

LOCUS BC039449

DEFINITION Homo sapiens hypothetical protein LOC339287, mRNA (cdna clone

ACCESSION IMAGS:5311571), partial cds.

VERSION BC039449

KEYWORDS BC039449.1 GI:25123321

SOURCE

ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 1269)

AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,J.S., Shenmen,C.M., Schlier,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Wax,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marsina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Utsid,T.B., Toshlyuk,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mulhally,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huiyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketterman,M., Madan,A.C., Shcherchenko,Y., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shcherchenko,Y., Bouffard,G.G., Blakeley,R.W., Touchman,D.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.B., Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 1269)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (01-NOV-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

COMMENT

Contact: MGC help desk

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshlyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIND)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
Series: IRAK Plate: 77 Row: n Column: 1.

FEATURES
source
1. 1269
Location/Qualifiers

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5311571"
/tissue_type="Brain, hypothalamus"
/clone_lib="NIH MGC_96"
/lab_host="DH10B"
/note="Vector: pBluescript"

ORIGIN

Alignment Scores:

Pred. No.: 4,13e-23 Length: 1269
Score: 1030.50 Matches: 200
Percent Similarity: 93.49% Conservative: 1
Best Local Similarity: 93.02% Mismatches: 7
Query Match: 31.83% Indels: 7
DB: Gaps: 2

US-10-054-935-2 (1-614) x BC039449 (1-1269)

QY 264 MetGluArgArgMetGluLeuValLysLysAspAsnGluLysGluArgHisLysLeuPhe 283
Db 1 ATGGAAAGGGGAGGAGCTGTAAGAGATACAGAAAGAAAGGACACAGCTGTT 60
QY 284 GlnGlyTyrGluThrGluGluArgGluGluThrGluLeuSerGluLysIleLysLeuGlu 303
Db 61 CAGGCGTATGAACTGAAGAGAGAGAGAAACAGAGCTATCTGAGAAATTAACGTGAG 120
QY 304 CysGlnProGluLeuSerGluThrSerGlnThrLeuProProlProPheSerCysGly 323
Db 121 TGCAGGAGGAGCTTCCGAGACATCCAGACTGCTGCCAAGCCCTTCATGTGGG 180
QY 324 ArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArgLysThrPro 343
Db 181 CGAGGTGAAAGGACATTAAGAAATCCCATTTGGAGTACGAAAGAAAGAACTCT 240
QY 344 ValLysLysLeuAlaProGluPheSerLysValLysThrLysProLysHisSerPro 363
Db 241 GTTAAAGCTGCTCTCGAATTTTCAAAAGTCAAAACAAACCTCTTAAGCCTCTCT 300
QY 364 IleLysGluGluProCysGlySerLeuSerGluThrValCysLysArgGluLeuArgSer 383
Db 301 ATTAAAGAGAACCTGTGTTCTTATCTGAAACGTGTTGTAAACGTGAATTGAGAGC 360
QY 384 GlnGluThrProGluLysProArgSerSerValAspThrProProlArgLeuSerThrPro 403
Db 361 CAAAGAACCCCAAGAAAGCCCGGCTCTTCACTGACACCCCAAGACCTCTCCATCCC 420
QY 404 GlnLysGlyProSerThrHisProlLysGluLysAlaPheSerSerGluIleGluAspLeu 423
Db 421 CAAAGAGGAGCCACAGACCATCCCAAGAGAAAGCTTCTCAATGAGATGAAGATTGG 480
QY 424 ProTyrLeuSerThrThrGluMetTyrLeuLysArgThrPheGlnProProlProSerPro 443
Db 481 CCGTACCTTTCCACACAGAAATGATTTGTGTGGCCACAGCTCCCTCCATCACCG 540
QY 444 LeuProLeuArgGluSerSerProLysLysGluGluGluThrValAlaArgCysLeuMetPro 463
Db 541 TTACCATTAAGGGAATCTCTCCAAAGAGAGAGAGACTGTAGCA----- 585
QY 464 SerSerValAlaGlyGluThrSerValLeuAlaValProSerTrp 478
Db 586 ---AGTAAGCATAGAGAACACTGCTCTT---ATACCTTAGTGG 624
RESULT 12
LOCUS BV178745 1092 bp DNA linear STS 10-JUN-2004
DEFINITION Bgsm102348 Human DNA (Sequenc) Homo sapiens STS genomic, sequence

tagged site.

ACCESSION BV178745
VERSION BV178745.1 GI:48015232
KEYWORDS STS.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 1092)

AUTHORS

Nelson, R.M., Marxell, G., Kammerer, S., Hoyal, C.R., Shi, M.M.,

TITLE

Large-Scale Validation of Single Nucleotide Polymorphisms in Gene

JOURNAL

Genome Res. (2004) In press

COMMENT

Contact: Andreas Braum
Pharmaceuticals division
Sequenc, Inc.
3595 John Hopkins Court, San Diego, CA 92121, USA
Tel: 18582029018
Fax: 18582029020
Email: abraum@sequenc.com
Primer A: No primer sequence submitted
Primer B: No primer sequence submitted
STS size: 1092.

FEATURES
source

1. 1092
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="Human DNA (Sequenc)"
<1. .>1092

ORIGIN

STS

Alignment Scores:

Pred. No.: 7.48e-23 Length: 1092
Score: 1020.00 Matches: 219
Percent Similarity: 76.01% Conservative: 25
Best Local Similarity: 68.22% Mismatches: 58
Query Match: 31.50% Indels: 20
DB: Gaps: 8

US-10-054-935-2 (1-614) x BV178745 (1-1092)

QY 11 AlaAlaAlaProAlaGlyLysProlGluAlaArgLeuAspTyrGluArgAlaAla 30
Db 2 GCGCGGCGCCCTGCGCGGCAATCTGAGCAGGACTGACGAGCGGCGCGCG 61
QY 31 LeuGlyLysProGluAspGluProGlyAlaAlaGlyAlaHisPheLeuProArgHisArg 50
Db 62 CTGGGCGGCGCCAGAGACGAGCTGGGCGGCGGCGGAGCCCACTTCTCCCGGACCT 121
QY 51 LysLeuLysGluProGlyProProlLeuAlaSerSerGlnGlyLysSerProAlaProSer 70
Db 122 AAGCTCAAGAGAGCGGCGGCGGCGGCTGCTCTCCAGAGGCGGAGCGGCGGCTTCC 181
QY 71 ProAlaGlyCysGlyLysGlyLysArgGlyLeuLeuLeuProAlaGlyAlaAlaProGly 90
Db 182 CCGGCGGCGCTGGCGGCGGCAAGGCGGCGGCTTGTACTCCCGGCGGCGGCGGCGG 241
QY 91 GlnGlnGluGluSerTrpGlyLysSerValProLeuProCysProlProProAlaThrLys 110
Db 242 CAGAGAGAGAGAGCTGGGCGGCTGGGCGGCTTGCCTTGCCTTGCCTTGCCTTGC 301
QY 111 GlnAlaGlyLysGlyLysGlyLysProAlaAlaAlaAlaGlyCysSerProArgProLys 130
Db 302 CAAAGCGGCAATTTGGGCGGAGAGCTGCGGAGCGGAGCGGAGCTTGCAGAGCGGCGG 361
QY 131 TyrGlnAlaValLeuProLysGlnThrGlySerLeuValAlaAlaLysGluThrThr 150
Db 362 TATCAGGCGGCTGCGGCAATTCAGACGCGGCTCTCTGCGGCGGCGGCGGCAAGAGCT 421
QY 151 ProTyrAlaGlyAspLysGlyLysAlaAlaSerProAlaAlaThrAlaSerProAla 170

[illegible]

	STS	/clone.lib="Human DNA (Sequenciom)" <1..>580
ORIGIN		
Alignment Scores:		
Pred. No.:	3.2e-22	Length: 580
Score:	991.00	Matches: 184
Percent Similarity:	97.37%	Conservative: 1
Best Local Similarity:	96.84%	Mismatch: 5
Query Match:	30.61%	Indels: 0
DB:	11	Gaps: 0
US-10-054-935-2 (1-614) x BVL67952 (1-580)		
OY	11 ALaAaAlaPProAlagIyGIyaMnProGluInArgLeuApTyTyrGuArGaAlaAla 30	
Db	2 GCCCGCGGCCCTGGCCGGCGCAATCTAGACAGAGCATGTGACTACGAGCGGGCGCGC 61	
OY	31 LeuGIyGIyPProGIyuAPGIuProGlyVAlAlaGlunLahIsphLeuPProARGHIsARG 50	
Db	62 CTGGCGCGGCCCGGAGAGAGAGCTGGGGCGGCCAGACCACATTCCTCCCCGCACCGT 121	
OY	51 LysLeuLyvGIuPProGIyPProProLeuAlAserSerGlnGIyGIySerProAlaProSer 70	
Db	122 AAAGCTCAAAGAGCGCGGGGCCCGCCGCTGCTCTCCAGAGGGCGGAGACCCCGCCTTCC 181	
OY	71 ProAlagIyCyvSIyGIyLYvGIyARGIyLeuLeuLeuProAlaGIyAlaIAProGIy 90	
Db	182 CCGCGCGGCTGGCGGCGGAAGGGCGGGGGCTGYTTACTCCGCGCGGGCGGCCCGCGG 241	
OY	91 GlnGlnGlnGlnUsertTPGIyGIyServalProleuProcySPProProAlathrls 110	
Db	242 CAGAGGAAAGAGAGCTGGGGGGGTTCGGTGCTTCGTCMTGTCCGCCCGGCACAGAAG 301	
OY	111 GluAlaGIyLIegIyGIyGluProAlAlaAlaGIyAlaGIyCyvSerProARyProlys 130	
Db	302 CAAGCCGCGCATTTGGGGGGGAGAGCTGGCGGAGCCGAGAGCCGGCTCACGCCCGCCAG 361	
OY	131 TyrGlnAlaValLeuProIIegInThrGIySerLeuValAlaAlaAlalyGluProThr 150	
Db	362 TATCAGGGGGGTGCTGCCCATTCAGACGGGCTCTCTCGTGGGGCGGGCCAAAGAGCTTAG 421	
OY	151 ProTrpaIagIyApPLyvgIyVAlAlaAserProAlaAlathrlAlaserAspProAla 170	
Db	422 CCTGGGGCTGAGGACAAGGGTGGGGCGGCCCTCCCCTGCACACGGCTCGGAGACCGAGC 481	
OY	171 GlyProPProleuPProleuPProGIyPProPProleuAlaPProPhraIathAlaGIy 190	
Db	482 GGACCCCGACCACTAACCTCTGCCCCGGGCGCCACCCCTCGCGCTCACCGGACCGCGGG 541	
OY	191 ThrLeuAlaAlaserGIuGIyARyTrPLys 200	
Db	542 ACCCTGGCGGCGACGAGGACGAGATTGGAGA 571	
RESULT 14		
LOCUS	AC148580	169993 bp DNA linear HTG 05-OCT-2004
DEFINITION	Gasterosteus aculeatus clone ch213-205p22, WORKING DRAFT SEQUENCE,	
ACCESSION	AC148580	
VERSION	AC148580.10 GI:53793740	
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.	
SOURCE	Gasterosteus aculeatus (three spined stickleback)	
ORGANISM	Gasterosteus aculeatus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
TITLE	Acanthomorpha; Acanthopterygii; Perciformes; Gasterosteidae;	
JOURNAL	Gasterosteidae; Gasterosteus.	
REFERENCE	1 (bases 1 to 169993) Ou C., Lin S., Postlethwait J. and Roe B.A. Gasterosteus aculeatus BAC Clone ch213-205p22 Unpublished 2 (bases 1 to 169993)	

QY	443	oleuProleuAaArgIuSerSerProlySylSgIuInuThValAlAaRyCyleuMePr	463
Db	77648	---CCTCTCGTGAAGACTTCCCTCTAAAGAAAGAGGTGGCCAGTGAAGTGACTCC	777030
QY	463	oSerSerValAlaGly-----	468
Db	77704	TCATGTAGTACATCATATGCTGATGCTTTTGAACCGGACCTATATCTTTTTCATCATTT	77763
QY	469	-----GluThSerVal-L	473
Db	77764	TAATTCATTGAATGTCCCAAGCGCTCAGTCTGAAATTACTTGATTCGCGCTCGCTCC	77823
QY	473	euaIaValAProSerThrAaRyAspHisSerValGluProleuAaGAPProAaProSerA	493
Db	77824	TCTTAGTTCCTTCATGAGAGAAATCCATCGACCTCG--GACGAGAACCCGCTCT	77880
QY	493	spleu-----	494
Db	77881	TTATTTCCCCCGAGGTGAGCTGCATCTGAAATGTAATACTGCCCCCGAAAGTTAT	77940
QY	495	-----LeuGluSerLeuAaAspSerValPheS	504
Db	77941	CTCAGTGAATPACTGACGTGAATGACTTCCCGTTAAAGCCGCTGATGACACAGTGTCT	78000
QY	504	eryIaAaRyHisAlaIyLeuGluLeuAaRyGluIySaTaGlyAs-----	518
Db	78001	TGAAGCCCACTGMAAGCTGAGTTGGACGAGAAAGAGAAAGATTGATGCTTTT	78060
QY	518	-----	518
Db	78061	GATTCMAAATCTCATTAATATATCATCTATCTGCGCTGATTTTGTTCATTAATTA	78120
QY	518	-----	518
Db	78121	AGTTGTTTTGTCAACATGTTATTTGAGTCAAGATTCACTTCTTAGTACGGAAGCTCA	78180
QY	519	-----ArgT	520
Db	78181	GCAGTAAATGAGACATGACTGAGCTGCTTTTACATTTGCTCCCTGTGTCTTCAGGT	78240
QY	520	rPaPpIleGluAaRyIleAaRyGluInuAaRyIleuGluAaRyLeuGluInuAaRyMetTyrL	540
Db	78241	GGGACATTCACAGATCAGAGACAGAGCGCATGTTTCAGCGTCTTCAGACAGGATGAACA	78300
QY	540	ysIySylSgIyIleGluInuSerGluProGluValThrSerPhePheProGluProAspA	560
Db	78301	GAAAGAGAGTGTCTCCGGAACGAGCGGAGCTGTCTCTTACCCAGACATGAAAG	78360
QY	560	sp-----LeuGluSerLeuAaAspSerValPheS	560
Db	78361	ACGGTAGCGTCTCATGCTCTCCACGMAAATGACAGACAGATGATGAATGCTGTTTTT	78420
QY	561	-----ValGluSerLeuMetIleThrProPheLeuProV	572
Db	78421	AAAACTCAATTCATTTATTTCCGCAATGTAAACAGATTGATCATCTCTTCTTCCCG	78480
QY	572	aIValaIaPheGlyAaRyProleuAaRyProlySleuThrProGlu-----	585
Db	78481	TGTTTGATTCGGCGCGCGCGTGTGCCAAAGTCTCAACAACAGTAAGTTCATTTTAAATC	78540
QY	585	-----	585
Db	78541	TGTTCTTGCCATTTTACTGTGTATGATGTTTGCTCATTTAGCATGTCATACAGTTATTT	78600
QY	585	-----	585
Db	78601	GGCATTGATGTATCTAATTTATATATATATATATATATATATATATATATATATATA	78660
QY	585	-----	585
Db	78661	TATATATATATTTTACCAATCTAGTAAAGAAAGAAATGAGAGATGGGAGAGAGTGAAG	78720
QY	585	-----	585

Db	78721	ATTATTTGGCTGAAGCACATTTGTAAGACTTGTGTACTGGTTTGGTAATTAGTACATT	78780
Qy	585	-----	585
Db	78781	TTATTACAGCGCTGGAGTTTAATTACAGTCTGTGACGTGATTAATAAATAATACATAACATTA	78840
Qy	586	-----	586
Db	78841	AGTGTTTTAAATACATGAACATTGATGATTTTCATTCTTCGACGAACTTCGAGTTGCCCT	78900
Qy	591	rpLeuaspGluArgSerSerCysArgLengluIleGlnIlyslYsgIlnThrProHisArgT	611
Db	78901	GGCTTGACGACGAGCGAAGCGCTGTTCGATCGAGGTGCCCAAAAACACACTCTCCACCGGA	79600
Qy	611	hrCysArgLys 614	
Db	78961	CGTCTCGGAAG 78971	
RESULT 15			
CQ842718		3425 bp	DNA
LOCUS	CQ842718		linear
DEFINITION	Sequence 1365 from Patent EP1440981.		PAT 02-AUG-2004
ACCESSION	CQ842718		
VERSION	CQ842718.1		GI:50894505
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J., Isono, Y., Nagai, K. and Irie, R.		
JOURNAL	Full-length human cdna		
FEATURES	Patent: EP 1440981-A 1365 28-JUL-2004; Research Association for Biotechnology (JP)		
source	Location/Qualifiers		
	1..3425		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
Alignment Scores:			
Pred. No.:	1,92e-16	Length:	3425
Score:	819.00	Matches:	155
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	25.29%	Indels:	0
DB:	6	Gaps:	0
US-10-054-935-2 (1-614) x CQ842718 (1-3425)			
Qy	460	CysLeuMetProSerSerValAlaGlyIuThrSerValLeuAlaValProSerTArg	479
Db	540	TGTCGTGATCCATCAAGGTGTGACGAGAACTTCAGTCTGTGCTGTCTCTTCCTTGGAGG	599
Qy	480	AspHisSerValGluProLeuArgAspProAsnProSerAspLeuLeuGlnAsnLeuAsp	499
Db	600	GACACCTCAGTACAGGCTCTTAAGGACCCAAATCTTCAACCTTTTGGAGAACTCGAT	659
Qy	500	AspSerValPheSerIlySarGHisAlaIlyLeuGlnLeuAspGluIlySarArgIlySarG	519
Db	660	GACAGTGTGTTTTCGAGCGCATGCAAACTGAGCTGATGAGAAAGAGAAAGAA	719
Qy	520	TrpAspIleGlnArgIleArgGluGlnArgIleLeuGlnArgLeuGlnLeuArgMetIlyr	539
Db	720	TGGATATTTTCAGAGAACTCAGGAAACAAAGATTTTTCAGCGACTGACGCTCAGAAATGTAT	779
Qy	540	IlyslYslYsgIlyIleGlnIlySerGluProGluIlyThrSerPhePheProGluProAsp	559
Db	780	AAAAAGAAAGAAATTCAGAAATCTGAGCTCGAGGTTTACCTCAATTTTTCCTGAGCCAGAT	839

Qy	560	AspValGluSerLeuMetIleThrProPheLeuProValValAlaPheGlyArgProLeu	579
Db	840	GATGTTGAAAGTTTGATGATTACCCCTTCTTGCTGTGTGACATTGGACGACCATTA	899
Qy	580	ProlyseLeuThrProGlnAsnPheGluLeuProTyrLeuAspGluArgSerArgCysArg	599
Db	900	CCAAATTAAGTCCACAGAAATTTGAGCTACCTGTGTGATGAGCGTAGCGGATGCAGA	959
Qy	600	LeuGluIleGlnLysLysGlnThrProHisArgThrCysArgLys	614
Db	960	TTGGAGATCCAGAGAGAACACACTCACCGGACGTGTAGAAA	1004

Search completed: March 26, 2005, 01:17:57
 Job time : 7851 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2005, 15:23:27 ; Search time 43 Seconds
(without alignments)
1065.919 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 3238
Sequence: 1 MTRKSAVFRAAAGAGNPE.....RSRCRLLEIKQKTPHRTCRK 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues
Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTCUS COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1a1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	202.5	6.3	8991	4 US-08-714-741-32	Sequence 32, Appl
2	199.5	6.2	905	2 US-08-574-959A-9	Sequence 9, Appl
3	199.5	6.2	905	3 US-09-357-014-9	Sequence 9, Appl
4	199.5	6.2	1135	2 US-08-574-959A-7	Sequence 7, Appl
5	199.5	6.2	1135	3 US-09-357-014-7	Sequence 7, Appl
6	199	6.1	1020	4 US-09-538-092-911	Sequence 911, Appl
7	187	5.8	712	4 US-09-949-016-10600	Sequence 10600, A
8	187	5.8	919	4 US-09-949-016-6954	Sequence 6954, Ap
9	186.5	5.8	904	4 US-09-976-594-615	Sequence 615, Appl
10	185	5.7	1325	4 US-09-294-298A-6	Sequence 6, Appl
11	185	5.7	1376	4 US-08-294-298A-4	Sequence 4, Appl
12	185	5.7	1423	4 US-09-294-298A-2	Sequence 2, Appl
13	176	5.4	1659	4 US-09-949-016-9752	Sequence 9752, Ap
14	174	5.4	527	4 US-09-370-838-216	Sequence 216, Appl
15	174	5.4	527	4 US-09-854-133-216	Sequence 216, Appl
16	173	5.3	1384	3 US-08-976-355-11	Sequence 11, Appl
17	172.5	5.3	707	4 US-09-919-039-278	Sequence 278, Appl
18	172.5	5.3	707	4 US-09-538-092-993	Sequence 993, Appl
19	172.5	5.3	735	4 US-09-949-016-10120	Sequence 10120, A
20	172	5.3	3969	4 US-09-538-092-1262	Sequence 1262, Ap
21	171	5.3	543	4 US-09-535-008-63	Sequence 63, Appl
22	171	5.3	577	4 US-09-535-008-61	Sequence 61, Appl
23	171	5.3	820	4 US-09-252-991A-23346	Sequence 23346, A
24	171	5.3	1646	4 US-09-535-008-67	Sequence 67, Appl
25	171	5.3	1647	4 US-09-535-008-2	Sequence 2, Appl
26	171	5.3	1647	4 US-09-824-574-4	Sequence 4, Appl
27	171	5.3	1647	4 US-09-538-092-1172	Sequence 1172, Ap

28	171	5.3	1649	4 US-09-535-008-75	Sequence 75, Appl
29	171	5.3	1650	4 US-09-535-008-71	Sequence 71, Appl
30	171	5.3	1678	4 US-09-535-008-69	Sequence 69, Appl
31	171	5.3	1679	4 US-09-535-008-67	Sequence 67, Appl
32	171	5.3	1681	4 US-09-535-008-77	Sequence 77, Appl
33	171	5.3	1682	4 US-09-535-008-73	Sequence 73, Appl
34	170.5	5.3	1665	4 US-09-538-092-1142	Sequence 1142, Ap
35	170.5	5.3	1665	4 US-10-274-978-2	Sequence 2, Appl
36	170.5	5.3	1665	4 US-10-274-978-2	Sequence 2, Appl
37	169.5	5.2	1034	4 US-09-252-991A-28921	Sequence 28921, A
38	169.5	5.2	3969	3 US-08-061-376-5	Sequence 5, Appl
39	167	5.2	2142	4 US-09-538-092-1142	Sequence 1142, Ap
40	166	5.1	2220	4 US-09-335-011-1	Sequence 1, Appl
41	165.5	5.1	802	4 US-09-823-240A-2	Sequence 2, Appl
42	164.5	5.1	758	4 US-09-904-987-5	Sequence 5, Appl
43	164.5	5.1	3256	4 US-09-919-172-98	Sequence 98, Appl
44	164.5	5.1	3256	4 US-09-976-594-22	Sequence 22, Appl
45	164.5	5.1	3256	4 US-09-919-039-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
GENERAL INFORMATION:
APPLICANT: Brlies, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swiatlo, Edwin
APPLICANT: Yocher, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tate, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
TITLE OF INVENTION: PORTIONS AND PRODUCTS
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtiss, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer Bsq., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 8991 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-714-741-32
Query Match 6.3%; Score 202.5; DB 4; Length 8991;

	Best Local Similarity	21.3%, Pred. No. 0.00027,	Mismatches 134;	Conservative 79;	Mismatches 224;	Indels 191;	Gaps 30.
QY	40	A A E A H P L P R H R K L E K P G P P L - - A S S Q G S P A P S P C G G K G R G L I L P A G A A P G Q O E S W G	97				
Db	6166	A A E E D L I A K K X K L T E K E A E A D L K K A V A D E P E T P A P A P A - - - - - P A - P A P A P P E A - -	6211				
QY	98	G S V L P C P P P A T T Q A G I G G E P P A A A G S G S P P K Y Q A V L P I O T G S L V A A K E P T P M A G D N G	157				
Db	6212	- - - P A P A P A P K - - - - - P A P A - - - - - P P X - - - - - A P A P K A P - - - - -	6238				
QY	158	G A A S P A A T A S D P A - G P P L P L P C G P P L A P T A G T L A A S E G R W K S M R S P G G G G S G A S	216				
Db	6237	- A P R P A - - - - - P A K P A P A P A P A P P E K P A L - - - - - K E I D S D V - - - - -	6270				
QY	217	S Q A A C L N Q I L L L Q D L I B Q O O O L O A K E I E E L K S E R D T L A R I E R M - - - - -	266				
Db	6271	- - - - - E V K A E I E L V K E A K E A P R N E K - V Q A A F A V E S K K A E A T R L E K I K T D R K A E	6322				
QY	267	- R M Q L V K D N E K E R - - - - - H K L P Q G Y E T E E R E B E T L S K I L E C Q P E L S E T S Q I L P P K P F	320				
Db	6323	A K R R A A E D K V K E K P A P K P E N P A B O P K A E K A D Q A E D Y A R S E E Y X R L T Q O O P P K T E	6382				
QY	321	S C G R S G - - - - - K G H R K S - - - - - P F A S T R K T P V K L A P E F S V K T - - - - - K T	358				
Db	6383	K P A P S T P K T G E A R E S K A E K V N O P K K E V S K - - K K E A T R L E K I K T D R K K A E A A K A	6439				
QY	359	P K H S P I K E P C G S L S E T V C K E L S Q E T P E K R S S V D P P R L S T P O K G S T H P E K A F S S	418				
Db	6440	A E E D K V K E K P A E Q P A P A P Q - - - - - P E K P A P A - - - - - P K E N P A E Q P K A - - E K P A D Q	6486				
QY	419	E I E D L P V I S T E M Y L C R M H O P P P E P L P R E S S P K K E E F V A A C L M P S S V A G E T S V L A P S W	478				
Db	6487	A E E D Y A R S E E Y X R L T Q O O P P K T E K A P O S T X I K E - - - - - X D S E X S - - - - -	6529				
QY	479	R D H S V E P L R D N P S D L - - - - - L E N L D S V - - F S K R H A K L E D E R R K - - - - -	518				
Db	6530	E D Y L K E G L R A P L O S L D T Y K A K L S K L E B L S D K I D E L A I N K L E V Q L D A G N N V E A Y F	6589				
QY	519	R M D I O R I E R C I L O R L Q L M Y K K G I O E S E - - - - - P E V T S F P E P D V - E S I M I	566				
Db	6590	K E G E K T T A E K K A E L E K A E A D L K K A V D E P I P P A P A P Q A P A P E K A E K P A P A P A P E K A P	6649				
QY	567	T P P L P V A F G R P L P K L P O N F E L P W D E	594				
Db	6650	A P E K P A P P E K P A P - - T P E T P K T G W K C E	6675				
RESULT 2							
US-08-574-959A-9							
; Sequence 9, Application US/08574959A							
; Patent No. 5962224							
GENERAL INFORMATION:							
APPLICANT: Jaekyoon Shin, Insil Jung, Ratna K. Vadlamudi							
APPLICANT: and Jack U. Strominger							
TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES							
TITLE OF INVENTION: AND USES THEREFOR							
NUMBER OF SEQUENCES: 22							
CORRESPONDENCE ADDRESS:							
ADDRESSEE: LAHIVE & COCKFIELD							
STREET: 60 Scate Street, Suite 510							
CITY: Boston							
STATE: Massachusetts							
COUNTRY: USA							
ZIP: 02109-1875							
COMPUTER READABLE FORM:							
MEDIUM TYPE: Floppy disk							
COMPUTER: IBM PC compatible							
OPERATING SYSTEM: PC-DOS/MS-DOS							
SOFTWARE: PatentIn							

```

ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION: DEN
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 905 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-574-959A-9

Query Match          6.2%; Score 199.5; DB 2; Length 905;
Best Local Similarity 21.9%; Pred. No.2,5e-05;
Matches 120; Conservative 53; Mismatches 225; Indels 149; Gaps 22

OY      38 PGAAEAHFLPHNRK-----LKEBGPPLASQGGSPAPSGCGGKRGILLPAGAAPG 90
DB      414 PTPAPVPLRLRHRSRSGPHRSILRAPCPQMA-----PCPOQAPCSAG-----PWPSAGP 462

OY      91 QQESWVG-----SVPLPCPP-----PATKOAGICGEPMAAGACS-----126
DB      463 VPSEBPWISTTANLIGLSRSPVCPRLLPGENHNHAGSNEDPI LAPSGTPPTTIPDETET 522

OY      127 ---PRPKY-----QAVLPITQGSLVAAAKE-----PTPWADKG 157
DB      523 GGRVPRAFAHYHDKESAADVAILSESDSDSVIYPEGLPPLPPPPSGATPTPIAPTPG 582

OY      158 GAASPATATSD----PAGPPLPLPGPPP--LAPTATAGTLAABGRMKSMRKSPLAG 209
DB      583 PTASPPVPAKEBPBELPAAPGPLPLPPPPPPPVDPVXLPPPOLVPEG-----TP--G 633

OY      210 GGSGGASSQAACLKQIILLDLEQQQQQLQAKKEITEELKSERDTLLARIEMERMQ 269
DB      634 GGGPPALAE-----DLTVININSSDEEEEGEHEEEEEEEEEEBE-----EEEEBE 682

OY      270 LVKKDNKERHKLFOGYETE-----ERRETELSEKIKLECOPELSETSQTLPPKPFS 321
DB      683 EDFEEBEDDEHYEYEEEEEHEEFEEFEHEEGELEEEEBEDEEBELEEVE---EDLE 739

OY      322 CGRSKGKHRRSPFGSTERKTPTVKLADEFKVKTKTPKSHSPIKEBPGCSLSEIVCKREL 381
DB      740 FGTAG-----GEVEEGAPPPPTLPALPPESPSPKVQPEPEPEBGLLEV-----784

OY      382 RSQETPEKRPSVUTPRRLS-----TPKGGSTHPHKKAFFSEIEDLYL 426
DB      785 --EEFGTEERGCADTAFTLAEALPSQGEVEREGESPAGP--PROELVEEPSXPPTL 839

OY      427 STTEMVLCRMHQPPSPPLPLERSSPKKETVARCI MPASVAGETSILA VPSWRDSHSEVPL 486
DB      840 LEEETEDGSDKVQPPETPAEE--METETAEALQEKEQDDTAAML--DFIDCPPDE 895

OY      487 RDPNPSPD 493
DB      896 KRPPEPTE 902

RESULT 3
US-09-357-014-9
; Sequence 9, Application US/09357014
; Patent No. 6291645
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Inseil Young, Ratna K. Vadlumludi
; and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510

```



```

RESULT 7
US-09-949-016-10600
: Sequence 10600, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CL001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: PRIOR FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10600
: LENGTH: 712
: TYPE: PRT
: ORGANISM: Human
US-09-949-016-10600

```

```

RESULT 8
US-09-949-016-6954
; Sequence 6954, Application US/09949016
; Patent No. 6812319
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 6954
; LENGTH: 919
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6954

```

Query Match	5.8#:	Score 187:	DB 4:	Length 919:
Best Local Similarity	21.5#:	Pred. No. 0.00018:		
Matches 106:	Conservative	75:	Mismatches 201:	Indels 110: Gaps 22:
Qy	101	PLP---CPPATKQAGIG-GEPAAGAAGCSBPXYOAVLPYOTGSLVAAAKEPTPMWADK	156	
Db	191	PLPRTLSPTASAVAPAPSQGIPTDSDEBTPKKSARILIESITVSLWATPDQPKQGVGT	250	
Qy	157	GGASPAATASDPKGPPLP-LPGPP-----PLAPATAGTLAASGRKMSRKSPLG	208	
Db	251	GRSASKRIAAVSPBGRDPSAPFPPSPMRERYLAFLPD-NESTGTGSTDQSARHSPIA	309	
Qy	209	GGGSGASSQAACLKOTLLQDLTLEBQQQOLAKKEKIEELKSEBDTLARI---ERME	265	
Db	310	PSPS-----POLYAKQTSLVAKQESVVRASRLAKTAEPPAASRIILCHSTLE	360	
Qy	266	RRMOLVKKDKNEKHKLFGQYETEEREETELSEKIKLECOBELSETGTLPPKPSCGRS	325	
Db	361	RLANV-----EVDQKVGSSQKEPPEABPV-AAABEPVENDN-----	396	
Qy	326	GKGHKRSPFSGSTERKTPVKKLAEFSGVVKTKTPKHSPKIEKPPGSLSETVCKELBSQE	385	
Db	397	-GNNSGPHNDT-----IANSTENPKPAASP-----E	423	
Qy	386	TPKPRRSVDTPRLSTSPQKGPSTHPKEKASV---SEIEDLSPYLSTEMYLCRMHOPPS	442	
Db	424	TPSAGQGEAKT-DQADGRPRPGARRKRSYKQAVSELDESHLEDEL-----QPRRS	476	
Qy	443	PLPIRESPPKEETVARCLMP-SSVAGETSVLAVPMSRHSV-----EPLR-DPNPS	492	

```
Db      477 KFP---SSCPASKVRLPFLTVORNOMLTPTSAFVSVKSPITKNTPLMDPKK 533
Qy      493 D--LENT--DVSFSKHALELDEKRRKMDIQIRIQRLIQRLQRLMYKKKGIQES 548
Db      534 ERORENTLRKEBAEQLRQKVE-EDKRRRLBEVKLRBEERLKVLAQARVERVQMEKK 592
Qy      549 PEVTSFPFPDD 560
Db      593 KOIEKFAQIDE 604
```

RESULT 9

```
US-09-976-594-615
; Sequence 615, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 615
; LENGTH: 904
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 2789525CD1
US-09-976-594-615
```

Query Match 5.8%; Score 186.5; DB 4; Length 904;

Best Local Similarity 19.2%; Pred. No. 0.00019;

Matches 113; Conservative 64; Mismatches 174; Indels 237; Gaps 23;

```
Qy      9 KAAAPAGNPBQRLYER-----AALGGPEDE----- 37
Db      469 KGGMAAASVQQRQRYRRNOQSSSSSSSEDRPRKSHVKNQGEVRRRRHSRS 528
Qy      38 ---PGAABHFLPRHKLKEPPPLAASQSGSPAPSPACGGKRGILLPAGAARQOE 93
Db      529 ASBPFRKQKETSPRRRRSPSP-PTRRRRSPSP- 570
Qy      94 ESMGGSVPLPCPPATKQAGIGEPAAAGCSPPRYQAVLPQTGSLVAAKERPMA 153
Db      571 RT-----PTPPRRRT- 608
Qy      154 GDKGAAPATASDPAGPPPLPLPGPPLAPTATAGTLAASGRWMSKSPVGGGGS 213
Db      609 -----PPKRTASPPPPKRRASPPSP-----KRVSHSPPPKQSSP----- 647
Qy      214 GASSQAACLKQILLDLLEQQOOLQAKEKEIEELKSRDTLLARIERMERMDLVK 273
Db      648 -----VTKRSPSLSKGR-----KGSPPSSTEARSPQPNKR 681
Qy      274 DNEKERNKLFQGYETERRETELSEKIKLECQPLSTSTQTLPPKPPSCGSGKHKRS 333
Db      682 HSPSP-----PRAQTSSS--PPPRKASSSPQRRQS 713
Qy      334 PFGSTKTPVKKLAPFSKVTKTPKSPIKEPQCSLSEYCKRELKSOETPEKRRS 393
Db      714 PESTR--PIRV-----SRTEPKIKK--AASPQSVRRVSSSRVSGSPEPA 760
Qy      394 VDPRLSLP--OKGPTH-----PKKAFSSRIEDLPYLSTTEMYLCRMHOPPPSLPL 446
Db      761 AKKPRAPSPVQGSSTWNSPAVPVKAKS-----PTSPSP 799
Qy      447 RSPSP-----KKEETV---ARCLMSSVAGETSVLAV 475
```

```
Db      800 KUSDGEGGKKKKKKDKKKKKKKKKKKKKAAGAVAAAAAATPAIAAATTLA- 858
Qy      476 PSWRHVSVEPLDNP-----SDLENTDVSFSKHALELDEKRRK 518
Db      859 -----QESPVAAPFPKKTESAEADNLD---LEKHLREKALRSWK 897
```

RESULT 10

```
US-09-294-298A-6
; Sequence 6, Application US/09294298A
; Patent No. 6723838
; GENERAL INFORMATION:
; APPLICANT: HOGANIR, RICHARD L.
; APPLICANT: KIM, GUEBHAH
; TITLE OF INVENTION: SIGNAL TRANSDUCING SYNAPTIC MOLECULES AND USES THEREOF
; FILE REFERENCE: 48235/1699
; CURRENT APPLICATION NUMBER: US/09/294,298A
; PRIOR FILING DATE: 1999-04-19
; PRIOR APPLICATION NUMBER: 60/082,690
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082,717
; PRIOR FILING DATE: 1998-04-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1325
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: mammalian
US-09-294-298A-6
```

Query Match 5.7%; Score 185; DB 4; Length 1325;

Best Local Similarity 20.2%; Pred. No. 0.00039;

Matches 117; Conservative 62; Mismatches 203; Indels 196; Gaps 23;

```
Qy      17 GNEQRLD-----YERAAALGPEDEPGAABHFLP-----RHRKLERGP 57
Db      802 GVPAAQLRLPLSFQNLFLHMAADGPRPAGNGSGSGHPPSSNNNNNNHNGGEPGD 861
Qy      58 PLASSG-----GSPRP-----SPAGCGGKRGIL-----LIP 84
Db      862 TPAFPHGYSKEDLSTGVPRPAASTLHSHSYSDERFGSTDTFTRQLSLQNLQNLSP 921
Qy      85 AGAAPQOEBSWGSVPLPCPPATKQAGIGEPAAAGCSPPRYQAVLPQTGSLVA 141
Db      922 PQTTPR-----PAPSGGSGSGSGSGGGGGGPR-----PIQRKSSQ 962
Qy      142 -LVAAKEPTPMAGDKGAASPAATASDPAGPPPLPLPGPPLAPTATAGTLAASGRWK 200
Db      963 LTVSAQKRPSSGNL-----LQSPSYGAPARFQGSLSKSGISGSGS-- 1008
Qy      201 SMKSPVGGGSGS-----ASSQAACLKQILLDL-----TE 234
Db      1009 -----GGGGGGLKPSITKQHSQTPSTLPTMPASERTVAWVNMPIISADIESAHIE 1061
Qy      235 QQQQOLQAKEKEIEELKSRDTLLARIERMERMDLVKQNEKERNKLFQGYETERRE 294
Db      1062 REEYKLEYSKSDSR-----LDVKYEIEEISIKELHMSNKL--EYERRLLSQE 1114
Qy      295 ELSEKIKLECQPLSTSTQTLPPKPPSCGSGKHKRKPFGSTERTKPVKKLAPFSKV 354
Db      1115 EQSKILMQQALDEGSKRLRQ-----QVEKDSQIKSTIGRLMLV 1156
Qy      355 KTKTPHNSPIKEPQCSLSEYCKRELKSOETPEKRRSVDT-----PR-----RLSTPQ 404
Db      1157 EEEIARDNPRMAP-----LPEPKRLIDAORGSPRPVVOQTRVMPH 1199
Qy      405 KG-----PSTPR-----KKAFFSRIEDLPYLSTTEMYLCRMHOPPPSLPL 453
Db      1200 LGTAMPQPRPHRPGCRSQRTASSGTPTTSPSTIDLLPCADYPCBPVVTDLPLSAADP 1259
```

QY 454 EETVARCLMPSSVAGE---TSVLA--VPSWRDHSVEPL 486
Db 1260 GTS-----PTTRGRHQGTAGLSPLTPPWGTHLSTPL 1291

RESULT 11
US-09-294-298A-4
; Sequence 4, Application US/09294298A
; Patent No. 6723838
; GENERAL INFORMATION:
; APPLICANT: HUGANIR, RICHARD L.
; APPLICANT: KIM, GJESHAH
; TITLE OF INVENTION: SIGNAL TRANSDUCING SYNAPTIC MOLECULES AND USES THEREOF
; FILE REFERENCE: 48235/1699
; CURRENT APPLICATION NUMBER: US/09/294,298A
; PRIOR APPLICATION NUMBER: 60/082,690
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082,717
; PRIOR FILING DATE: 1998-04-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1376
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: mammalian
; OTHER INFORMATION: SYNGAP-B
US-09-294-298A-4

Query Match 5.7%; Score 185; DB 4; Length 1376;
Best Local Similarity 20.2%; Pred. No. 0.00041;

Matches 117; Conservative 62; Mismatches 203; Indels 196; Gaps 23;

QY 17 GNEBQRID-----YERAAALGPEDEPGAALFLP-----RHRLKEKPGP 57
Db 853 GVAQOQLRILPLSFQNPFLHMAADGPGPRAHGSSGSHGPPSHNNHHHHHNGGEPGD 912
QY 58 PLASSQC-----GSPAP-----SPAGCGKGRL-----LLP 84
Db 913 TFAPFHYSKSEDLSTGVPRPRAASILHSHSYDEFGPSTDFTRQLSLQDNLQHLSP 972
QY 85 AGAPGQOESWGSVPLPCPPRATKQAGIGERPAAGAGCSRPKYQAVLPLOTGS--- 141
Db 973 PQTITGQR-----PABSGCGSGSGSGSGGCGQPP-----PLQKGSQ 1013
QY 142 -LVAAKEPTPMAGDKGAASPAATASDPAGRPPLPLPGPPLAPATAGTLAASEGRWK 200
Db 1014 LTVSAQOKPRPSSGNL-----LQSPFSGPAPRPROQSLSKESIGSGGS-- 1059
QY 201 SMKRSPLGGGCGSG-----ASSQAACLKQILLQLDL-----IE 234
Db 1060 -----GGGGGGGLKPSITKQHSQTPSTLNPMPASERTVAWVSNPHLSADIESAHIE 1112
QY 235 QOOOQLOAKKEBELKSEBDTLARIERMERMOQYKDNKEKRNHLFGQYETEEREET 294
Db 1113 REEYKLEKYSKSWDESR-----LDRVKEYEEBILHSIKERLHMSNRKL-BEYERRLLSOE 1165
QY 295 ELSEKILCEQPLSETSQTLPRKPFSCGSGKGNKRRKSPGSTERKTPYKULAPFSKY 354
Db 1166 EQSKILMOYQALRQSEKRLRQ-----QVEKDSQIKSIIGRLMLV 1207
QY 355 KXTTPKHSPIKEBPCCSLSETVCKRLRSQETPEKPRSSVDL-----PR-----RLSTPQ 404
Db 1208 EEEELRDHPMAABP-----LPEPKKRLDLDAORGSPFWVQOTRYVMPH 1250
QY 405 KG-----PSTHP-----KEKAFSSEIEDLRYLSTTEMYLCRMHPSPPLPRESPPK 453
Db 1251 LGTAMPQPPHPHPGCGSQRTASSGTPTQTSPPSITDILPLCARYPPSPVTDLLPSAPDP 1310
QY 454 EETVARCLMPSSVAGE---TSVLA--VPSWRDHSVEPL 486

Db 1311 GTS-----PTTRGRHQGTAGLSPLTPPWGTHLSTPL 1342

RESULT 12
US-09-294-298A-2
; Sequence 2, Application US/09294298A
; Patent No. 6723838
; GENERAL INFORMATION:
; APPLICANT: HUGANIR, RICHARD L.
; APPLICANT: KIM, GJESHAH
; TITLE OF INVENTION: SIGNAL TRANSDUCING SYNAPTIC MOLECULES AND USES THEREOF
; FILE REFERENCE: 48235/1699
; CURRENT APPLICATION NUMBER: US/09/294,298A
; PRIOR APPLICATION NUMBER: 60/082,690
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082,717
; PRIOR FILING DATE: 1998-04-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1423
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: mammalian
; OTHER INFORMATION: SYNGAP-A
US-09-294-298A-2

Query Match 5.7%; Score 185; DB 4; Length 1423;
Best Local Similarity 20.2%; Pred. No. 0.00042;

Matches 117; Conservative 62; Mismatches 203; Indels 196; Gaps 23;

QY 17 GNEBQRID-----YERAAALGPEDEPGAALFLP-----RHRLKEKPGP 57
Db 900 GVAQOQLRILPLSFQNPFLHMAADGPGPRAHGSSGSHGPPSHNNHHHHHNGGEPGD 959
QY 58 PLASSQC-----GSPAP-----SPAGCGKGRL-----LLP 84
Db 960 TFAPFHYSKSEDLSTGVPRPRAASILHSHSYDEFGPSTDFTRQLSLQDNLQHLSP 1019
QY 85 AGAPGQOESWGSVPLPCPPRATKQAGIGERPAAGAGCSRPKYQAVLPLOTGS--- 141
Db 1020 PQTITGQR-----PABSGCGSGSGSGSGGCGQPP-----PLQKGSQ 1060
QY 142 -LVAAKEPTPMAGDKGAASPAATASDPAGRPPLPLPGPPLAPATAGTLAASEGRWK 200
Db 1061 LTVSAQOKPRPSSGNL-----LQSPFSGPAPRPROQSLSKESIGSGGS-- 1106
QY 201 SMKRSPLGGGCGSG-----ASSQAACLKQILLQLDL-----IE 234
Db 1107 -----GGGGGGGLKPSITKQHSQTPSTLNPMPASERTVAWVSNPHLSADIESAHIE 1159
QY 235 QOOOQLOAKKEBELKSEBDTLARIERMERMOQYKDNKEKRNHLFGQYETEEREET 294
Db 1160 REEYKLEKYSKSWDESR-----LDRVKEYEEBILHSIKERLHMSNRKL-BEYERRLLSOE 1212
QY 295 ELSEKILCEQPLSETSQTLPRKPFSCGSGKGNKRRKSPGSTERKTPYKULAPFSKY 354
Db 1213 EQSKILMOYQALRQSEKRLRQ-----QVEKDSQIKSIIGRLMLV 1254
QY 355 KXTTPKHSPIKEBPCCSLSETVCKRLRSQETPEKPRSSVDL-----PR-----RLSTPQ 404
Db 1255 EEEELRDHPMAABP-----LPEPKKRLDLDAORGSPFWVQOTRYVMPH 1297
QY 405 KG-----PSTHP-----KEKAFSSEIEDLRYLSTTEMYLCRMHPSPPLPRESPPK 453
Db 1298 LGTAMPQPPHPHPGCGSQRTASSGTPTQTSPPSITDILPLCARYPPSPVTDLLPSAPDP 1357
QY 454 EETVARCLMPSSVAGE---TSVLA--VPSWRDHSVEPL 486
Db 1358 GTS-----PTTRGRHQGTAGLSPLTPPWGTHLSTPL 1389

```
RESULT 13
US-09-949-016-9752
; Sequence 9752, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9752
; LENGTH: 1659
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9752

Query Match          5.4%; Score 176; DB 4; Length 1659;
Best Local Similarity 20.1%; Pred. No. 0.0021;
Matches 114; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

QY 51 XLKEPRPLASGCGSPAPSPACCGKKGKGLL--PAGAARPGQDESSWGSSVLPCCPPA 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12 KMTSPPLP---GTPRPSPSPGPPSPGAMLGSPSPGSAHSMWG---PSPPPSA 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 109 ---TKAGIGSEPR-----AAAGAGSPRPKYQAV--LPITGSLVAAKEPTP 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 65 GHPRIQGGGCTYQDMHQMNMKMEKMSDDPRYNOMKMGKMSGGHAGMPPSP 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 152 WAGDKGGAASPAATASDPAPRPPLPLPGPPPLAPTAAGTLAASEGRWKSRKSLPGGG 211
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 MDOHSGVSPSLGSSHNASPVPAAGSSGP-----QMSGGP--GAP 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 212 GSGASQAACLKILLLDLLEQOQOOLAKKEIIBELKSE--RDTLARIEMERRM 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 166 LDGADPQA-----LGOQNRGPTTPFNQNLHQLRAQIMAYKMLAGQLPDHLQ 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 270 LVKDKNEKERHKLFGQYETEBERETELSEKIKLECOPELSETQTLPPKPFSGRSGKG- 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 214 MA-----VQGRP-----MFGMOQMPILPPPSVSATGPGGP 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 329 -----HKRKSPPGSTERTKTPVKKLAPERSKVKTTPKHSPIKEBPCSLSE 374
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 247 GPGPGPGPAPPNVSRPHMGGMPPPPPGSGVPPGMPGQ---PFGPPKPPBEGPMAN 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 375 TVCKELRKSQETPEKRRSSVDTPRLSTPOKGPSTHPEKAFSSSEIDLPYLSTTEMYLC 424
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 304 AAAPTSPQKLTLPQCTGPPSPAPPAVPVPASVPMFPQOTPCQAPQAPAMVPL----- 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 435 RMHQPSPPLP-----LRSSPKKEETVARCU-----MPSVAGETSVLAVPS 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 358 --HOKGSRITTPICKPGLDVEILQEBEYFLQARIARIELENTLPGSLGDLTKYAT-- 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 478 WRDHSVEPLRDPNSDLLENTDVSFKRAKLE--LDERKRRKWDIQRITREORILLQRLQ 535
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 414 ---IELKALRLNFORQLRG--EVVVCRRDPTALETLNANAYKRSKRQSLREARITKLE 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 536 LRM---YKKKGIOSEBEVTSPFPEPDD 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 470 KOOKIEOKRRRQKHQBYLNSILQAHKD 497
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
```

```
US-09-370-838-216
; Sequence 216, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; PRIOR FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-216

Query Match          5.4%; Score 174; DB 4; Length 527;
Best Local Similarity 23.1%; Pred. No. 0.0069;
Matches 143; Conservative 44; Mismatches 186; Indels 246; Gaps 36;

QY 26 ERRA--AIGPDEPGAAL-AHFLPRHKLK-----EPGPPLASGCGSPAPSP 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8 EKAENGALGSEPEREKVLNGBELTPRREKALENGELRSPAGSKVLNGLTPPKSE 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 72 AGCGGGRLLLPAGA-----APGOE--ESMGSSVLP--PCP-----PP 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68 DKVSENG-GLRPRTNTERPETGPMWRAPPEWKTSPSPDAPLTIGEPABETSIERRAPDS 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 108 ATKAGIGSEPPAAAGAGSPRPKYQAVLP-----IQTSLVAAKEPTPMAG--DKG-G 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 127 AVSSNNGSETAPPLG--PARKNGTLEPOTERRAPETGAPRA-----PGAGRLDGG 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 AASPAATASDPAG-----PPPLPLPGPP----- 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 180 GRAPVGTGAPGGGPGSGVDAKAGWVDNTRPPPPPLPPPPAQRRLBPAPRAPREV 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 -----PLATPATAGTLAASBGRN--KSMRKSPLGGGGGSGASSQAACLKQILLDL 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 240 APEGEGAPDSRAGDPTALSGDGPCKPBRKGP-----EMRPLFLDLG 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 234 EQOQOOLQAKKEIIBELKSEBDTLARIEMERRMQLVKKDKNEKERHKLFGQYET- 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 283 PPOGNSQIYKAR-----LSRLSLAPPLTLTPPPGPGPPRPMEGADAGAAG 330
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 289 -----EEREETELSEKIKLECOPELSETQTLPPKPFSGRSGKGHRRKSPF- 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 331 EAGGAGAPGAEBDEGDDEDEE-----EDEBAAPGA--AAGPGPGGARAAPVP 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 336 -----GSTERKTPVKKLAPERSKVKTTPKHSPIKEBPCGSLSETVCKELR- 383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 380 VVVSSADAAPARLRLGL-----LKSPPRGADDEPD-----SELRRKMSVFNHDTV 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 384 -----QSTPEK-----PRSSVD-----PPRLSTPOKG--PSTHPEKAFSSSEI- 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 YLPDQETPNELSVQAPPGDDTDPSTPPRAPPTPPHATFGDGFPS--NDSGFGSFBWA 483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 421 EDLPYLSTTEMYLCRMHQPSPPLPLRBSPPKKEETVARCLMPSVAGETSVLAVPSWRD 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 484 EDFPFL-----PPGP-PL-----CFSRFSV--SPALETG--- 511
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 481 HSYVEPLRDPN--PSDILEN 497
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 512 ---PPARADAPAGAVEN 527
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2005, 15:13:55 ; Search time 167 seconds
(without alignments)
1421.981 Million cell updates/sec

Title: US-10-054-935-2

Sequence: 1 MTRSAVFRAAAGAGNDE.....RSRCRLRIQKQTPHRTCKR 614

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: A_Geneseq_16Dec04:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3238	100.0	614	6	ABR58310
2	1446	44.7	281	3	AAB58317
3	1408.5	43.5	380	7	ADFS6888
4	1277	39.4	269	3	AAB42665
5	971	30.0	203	7	ADCC06840
6	897.5	27.7	189	7	ADCC06841
7	897.5	27.7	192	7	ADCC06839
8	806	24.9	153	8	ADCC06392
9	377.5	11.7	195	4	AAU30711
10	377.5	11.7	195	7	ADFC0190
11	264	8.2	60	8	ADP84563
12	227.5	7.0	1119	6	ABG75823
13	226.5	7.0	720	5	AAE20839
14	226.5	7.0	766	5	AAE20838
15	226.5	7.0	973	5	AAE20799
16	226.5	7.0	973	5	ABG64675
17	226.5	7.0	973	8	ADL77942
18	219.5	6.8	699	4	AAAM4452
19	215.5	6.7	536	8	ADM97710
20	215.5	6.7	1130	8	ADM97695
21	213	6.6	1038	4	ABR71916
22	209.5	6.5	777	5	ABP41839
23	209.5	6.5	1130	7	ADCC9065
24	209.5	6.5	1130	8	ABM81812
25	208.5	6.4	1251	7	ADE09332

26	208.5	6.4	1251	8	ADM87788	Adm87788 Human EST
27	206.5	6.4	905	2	AAW31186	Aaw31186 Human p16
28	206.5	6.4	983	8	ADM87342	Adm87342 Human p16
29	206.5	6.4	1135	2	AAW31185	Aaw31185 Human p16
30	206	6.4	1023	5	AAU82954	Aau82954 Human hom
31	205.5	6.3	676	4	AAAM3666	Aam3666 Human pol
32	205	6.3	615	8	ADQ36684	Adq36684 Trypanoso
33	203.5	6.3	2340	2	ADH11419	Adh11419 Human hu-
34	203.5	6.3	2365	7	ADD93763	Add93763 Human hel
35	203.5	6.3	2429	7	ADD93765	Add93765 Human hel
36	203.5	6.3	2432	3	AAV85565	Aay85565 Human hom
37	202.5	6.3	8991	6	ABU08487	Abu08487 S. pneumo
38	200.5	6.2	1021	7	ADJ68507	Adj68507 Human hea
39	200	6.2	1026	7	ADK11437	Adk11437 Human neu
40	200	6.2	1026	7	ADJ69309	Adj69309 Human hea
41	200	6.2	1026	8	ADQ19720	Adq19720 Human sof
42	199	6.1	617	4	AAAM6458	Aam6458 Peptide #
43	199	6.1	617	4	ABR35445	Abbr35445 Peptide #
44	199	6.1	617	4	ABR20884	Abb20884 Protein #
45	199	6.1	617	4	AAAM56269	Aam56269 Human bra

ALIGNMENTS

RESULT 1	
ABR58310	
ID	ABR58310 standard; protein; 614 AA.
AC	ABR58310;
DT	06-JUL-2003 (first entry)
DE	BCU1041 protein #SEQ ID 66.
OS	Homo sapiens.
PN	WO2003029421-A2.
PD	10-APR-2003.
PF	02-OCT-2002; 2002WO-US031287.
PR	03-OCT-2001; 2001US-0326526P.
PR	14-MAY-2002; 2002US-00144194.
PA	(ORIG-) ORIGENE TECHNOLOGIES INC.
PI	Sun Z, Li X, Fan W, Kovacs KF, Jay G;
DR	WPI; 2003-381623/36.
DR	N-FSDB; ACCT2044.
PT	New isolated human differentially-regulated breast cancer polynucleotide
PT	and polypeptide, useful for diagnosing, staging, prognosticating,
PT	preventing and/or treating diseases and conditions relating to breast
PT	cancer.
PS	Claim 6; SEQ ID NO 66; 127pp + Sequence Listing; English.
XX	
XX	The invention relates to isolated polynucleotides which are
XX	differentially-regulated in breast cancer. The methods and compositions
XX	of the present invention are useful for detecting, diagnosing, staging,
XX	monitoring, prognosticating, preventing and/or treating diseases and
XX	conditions relating to breast cancer, and may be used in gene therapy or
XX	antisense therapy. They can also be used in research, drug discovery,
XX	clinical medicine and forensic medicine. Sequences given in records
XX	ABR58278-ABR58323 and ABR58346-ABR58362 represent polypeptides encoded by
XX	polynucleotides of the invention that are differentially-regulated in
XX	breast cancer. NOTE: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pot_sequences
XX
SQ Sequence 614 AA;

Query Match	100.0%	Score 3238;	DB 6;	Length 614;
Similarity	100.0%	Prod. No. 1.1e-200;		
Best Local Match	0;	Mismatches 0;	Gaps 0;	
Best 614; Conservative	0;			

QY	I	MTMSAVAKKAAAAAPAGNPNDEQLDYERAAALGDPEDPEPAEAENHLPRHRKLKEPBPPLA	60
Db	1	MTMSAVAKKAAAAAPAGNPNDEQLDYERAAALGDPEDPEPAEAENHFLPRHRKLKEPBPPLA	60
QY	61	SSOGSPAPSPAGCCGGKGRLLPYGAAAGQQOESWGSVPLPCPPPATKQAGIGGEPPAA	120
Db	61	SSOGSPAPSPAGCCGGKGRLLPYGAAAGQQOESWGSVPLPCPPPATKQAGIGGEPPAA	120
QY	121	AGAGSPRPRTQAVLPIQTGSLVAAAKEPTPMAGDKGAASPAATSDPADGPPLPLPGP	180
Db	121	AGAGSPRPRTQAVLPIQTGSLVAAAKEPTPMAGDKGAASPAATSDPADGPPLPLPGP	180
QY	181	PPLAPTATAAGTLAAEEGRWKSRKSPDLGGGGSGASSQAACLKQILLDILTEGQQOOL	240
Db	181	PPLAPTATAAGTLAAEEGRWKSRKSPDLGGGGSGASSQAACLKQILLDILTEGQQOOL	240
QY	241	QAKEKEITELKSERDTLLARIERMERRMQLVKKDNEKERHKLFQGYETERRETELSEKI	300
Db	241	QAKEKEITELKSERDTLLARIERMERRMQLVKKDNEKERHKLFQGYETERRETELSEKI	300
QY	301	KLECPPELSETSQTLPPKPFSCGRSGKGHKRRKSPFESTERTKTPVKKLAEFSKVTKTTPK	360
Db	301	KLECPPELSETSQTLPPKPFSCGRSGKGHKRRKSPFESTERTKTPVKKLAEFSKVTKTTPK	360
QY	361	HSPIKEEPCGSLSEFTVCCKRELRSQETPEKRSSVDTPPRIISTPOKGPSTHPKKAFFSSEI	420
Db	361	HSPIKEEPCGSLSEFTVCCKRELRSQETPEKRSSVDTPPRIISTPOKGPSTHPKKAFFSSEI	420
QY	421	EDLPLLTSTTEMWLCTWHOPRPSPLPIRESSPKKEEVAACLMPSSVAGETSVALVSMD	480
Db	421	EDLPLLTSTTEMWLCTWHOPRPSPLPIRESSPKKEEVAACLMPSSVAGETSVALVSMD	480
QY	481	HSVEPLRDPNPEDLLENLDLSVFYSKHAALKLEDEKKRKMDIORIREORILQRLQRMKY	540
Db	481	HSVEPLRDPNPEDLLENLDLSVFYSKHAALKLEDEKKRKMDIORIREORILQRLQRMKY	540
QY	541	KKGIQSESPEVTSFPPEPDVESLMITPPLPVVAFGRPLPKLTPONFELPWIDERSCRCL	600
Db	541	KKGIQSESPEVTSFPPEPDVESLMITPPLPVVAFGRPLPKLTPONFELPWIDERSCRCL	600
QY	601	EIOKKQTPHRTCRK 614	
Db	601	EIOKKQTPHRTCRK 614	
RESULT 2			
ID	AAB58317	standard; protein; 281 AA.	
XX	AAB58317;		
XX	14-MAR-2001	(first entry)	
XX	Lung cancer associated polypeptide sequence SEQ ID 655.		
KM	Human; lung cancer associated protein; neuroprotective; cytoprotatic; cardioactive; immunomodulatory; muscular active; vulnerary; gastrontestinal; nephrotoxic; antilifeactive; gynecological; antibacterial; diagnosis; neural disorder; immune disorder; reproductive; proliferative disorder; wound healing; infectious disease.		
OS	Homo sapiens.		
FN	WO20055180-A2.		

XX 21-SEP-2000 .
 PD
 XX
 XX 08-MAR-2000; 2000WO-US005918 .
 PF
 XX
 XX 12-MAR-1999; 99US-0124270P .
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC .
 PA (ROSE/) ROSEN C A .
 XX
 XX
 PI Ruben SM;
 XX
 XX WPI; 2000-587514/55 .
 DR
 DR N-PSDB; AAF18193 .
 XX
 XX
 XX Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer .
 PT
 XX
 XX Claim 11; Page 1153-1154; 1425pp; English .
 PS

CC Polynucleotide sequences AA1719982 - AA178424 encode human lung cancer
CC associated proteins represented in AAB55106 - AAB58549. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnery; gastrointestinal
CC general; nephrotoxic; antineutic; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the protein
CC or polynucleotide sequences. The lung cancer associated polynucleotide
CC sequences may be used for detection of lung cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The proteins may be used to treat disorders such as
CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
CC cardiovascular, renal, and proliferative disorders. The proteins may also
CC be used in the treatment of wounds and infectious diseases.
CC Polynucleotide sequences AA181425 - AA181433 and peptide AAB58549 are
CC used in the course of the invention for the identification and
CC characterisation of the polynucleotide and protein sequences
XX
XX Sequence 281 AA;

Query Match	44.7%	Score 1446;	DB 3;	Length 281;
Best Local Similarity	100.0%	Pred. No. 2,7e-89;		
Matches 278; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

QY	181	PLPLPTATAGTGLTASBGRWKNKRSPLGGGGSGASSDAACLKQTLIIQLDPIEQOOOL	240
Db	1	PLPLPTATAGTGLTASBGRWKNKRSPLGGGGSGASSDAACLKQTLIIQLDPIEQOOOL	60
QY	241	QAKKEIIEELKSEBDTLIARIERNERRNQVYKONKERRHLFOGYETREBETELSEKI	300
Db	61	QAKKEIIEELKSEBDTLIARIERNERRNQVYKONKERRHLFOGYETREBETELSEKI	120

Qy	301	KLECPPELSETSQTLPPKPPSCGSGKGNKKRSPGSGNERKTQPVKKLADEPSKVYTKTPK	360
Qy	121 <td>KLECPPELSETSQTLPPKPPSCGSGKGNKKRSPGSGNERKTQPVKKLADEPSKVYTKTPK <td>180</td> </td>	KLECPPELSETSQTLPPKPPSCGSGKGNKKRSPGSGNERKTQPVKKLADEPSKVYTKTPK <td>180</td>	180
Db	361 <td>HSPIKEBPCGSLSETVCKRELRSQETPEKPPSSVDTPEPLSTLPQKGPSTHPEKAFSSEI <td>420</td> </td>	HSPIKEBPCGSLSETVCKRELRSQETPEKPPSSVDTPEPLSTLPQKGPSTHPEKAFSSEI <td>420</td>	420
Qy	181 <td>HSPIKEBPCGSLSETVCKRELRSQETPEKPPSSVDTPEPLSTLPQKGPSTHPEKAFSSEI <td>240</td> </td>	HSPIKEBPCGSLSETVCKRELRSQETPEKPPSSVDTPEPLSTLPQKGPSTHPEKAFSSEI <td>240</td>	240

[illegible]

XX DE Human polypeptide sequence SEQ ID NO:1096.
 XX KM biological activity; genetic engineering; hybridisation probe; oligomer;
 XX KM primer; chromosome mapping; gene mapping; recombinant protein production;
 XX KM human.
 XX OS Homo sapiens.
 XX PN MO2003080795-A2.
 XX XX
 XX XX 02-OCT-2003.
 XX PF 09-AUG-2002; 2002MO-US025485.
 XX PR 09-AUG-2001; 2001US-0311261P.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Yang Y, Wang Z, Weng G, Ma Y;
 XX DR N-PSDB; ADF57688.
 XX DR WPI: 2003-876918/81.
 XX PT New polynucleotides, useful as hybridization probes, oligomers or
 XX PT primers, for chromosome or gene mapping, for the recombinant production
 XX CC of proteins, and for generating antisense DNA or RNA.
 XX PS Claim 20; SEQ ID NO 1096; 571bp; English.
 XX CC The present sequence represents a polypeptide (II) with biological
 XX CC activity, which is encoded by an isolated polynucleotide sequence (I)
 XX CC from the present invention. Also described: (1) a vector comprising (I);
 XX CC (2) an expression vector comprising (I); (3) a host cell genetically
 XX CC engineered to comprise (I) which is operatively associated with a
 XX CC regulatory sequence that modulates expression of (I) in the host cell;
 XX CC (4) a polypeptide (II) encoded by (I); (5) a composition comprising the
 XX CC polypeptide of (4) and a carrier; (6) an antibody directed against the
 XX CC polypeptide of (4); (7) detecting (I) or the polypeptide of (4) in a
 XX CC sample; (8) identifying a compound that binds to the polypeptide of (4);
 XX CC (9) producing the polypeptide of (4); and (10) a collection of
 XX CC polynucleotides comprising at least one of the polynucleotide sequences
 XX CC (I). The polynucleotides (I) can be used as hybridisation probes,
 XX CC oligomers or primers, for chromosome or gene mapping, for the recombinant
 XX CC production of proteins, and for generating antisense DNA or RNA.
 XX SQ Sequence 380 AA;
 Query Match 43.5%; Score 1408.5; DB 7; Length 380;
 Best Local Similarity 80.8%; Pred. No. 1.3e-86;
 Matches 274; Conservative 4; Mismatches 10; Indels 51; Gaps 3;
 QY 311 TSQTLPP-----KPFSGRSG-----KGKRRSPF 335
 DB 58 STHTFPNNRVFAISSVATROP CSTGRGWEPGIIIVFSCOVPIAAKVNAPAFSWKSPF 117
 QY 336 GSTERKTPVKKLAPEEKKTKTKPKSPKKEEPCGSSSEVYCKRELASQETPEKRRSSVD 395
 DB 118 GSTERKTPVKKLAPEEKKTKTKPKSPKKEEPCGSSSEVYCKRELASQETPEKRRSSVD 177
 QY 396 TPPLRLSPQKGPSTHPKEKAFSSRIEDLPYLSTEWMLCRWHOPPPPLPLRESPPKEE 455
 DB 178 TPPLRLSPQKGPSTHPKEKAFSSRIEDLPYLSTEWMLCRWHOPPPPLPLRESPPKEE 237
 QY 456 TVARCLMPSSVAGETSVLAVPWSWDHSEVPLRDNPSDLLENLDVSFVKRAHAKLELDEK 515
 DB 238 TV-----AIPSWRDHSEVPELRDNPDDLLENLDVSFVKRAHAKLELDEK 281
 QY 516 RRRKWDIORIREQIILRIOLRMVKKKGIOSEBEVSPFPEPDVDSIMITPLPIVAVF 575
 DB 282 RRRKWDIORIREQIILRIOLRMVKKKGIOSEBEVSPFPEPDVDSIMITPLPIVAVF 341
 QY 576 GRPLPKLTPQNFELPWLDERSRCRLSIQKQTPHRTCRK 614

DB 342 GRPLPKLTPQNFELPWLDERSRCRLSIQKQTPHRTCRK 380
 RESULT 4
 AAB42665
 ID AAB42665 standard; protein; 269 AA.
 XX AAB42665;
 AC XX
 XX XX 08-FEB-2001 (first entry)
 DT XX
 XX XX Human ORFX ORF2429 polypeptide sequence SEQ ID NO:4858.
 DE XX
 XX XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 XX KM vulnerability; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 XX KM anticonvulsant; osteopathic; antiarthritic; immunosuppressive; cardiant;
 XX KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 XX KM hypotensive; dermatological; immunosuppressive; antihypertensive;
 XX KM antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 XX KM antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 XX KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 XX KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 XX KM cholesterol ester storage; systemic lupus erythematosus; infection;
 XX KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 XX KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 XX KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 XX KM thrombosis; contraceptive.
 XX OS Homo sapiens.
 XX XX
 XX XX WO200058473-A2.
 XX XX 05-OCT-2000.
 XX PF 31-MAR-2000; 2000MO-US008621.
 XX PR 31-MAR-1999; 99US-0127607P.
 XX PR 02-APR-1999; 99US-0127636P.
 XX PR 05-APR-1999; 99US-0127728P.
 XX PR 30-MAR-2000; 2000US-00540763.
 XX PA (CURA-) CURAGEN CORP.
 XX PI Shinketsu RA, Leach M;
 XX DR WPI: 2000-602362/57.
 XX DR N-PSDB; AAC76874.
 XX PT Novel nucleic acids and peptides derived from open reading frame X,
 XX PT useful for treating e.g. cancers, proliferative disorders,
 XX PT neurodegenerative disorders and cardiovascular disease.
 XX PS Claim 11; Page 4046; 5507bp; English.
 XX XX
 XX CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
 XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 XX CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
 XX CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
 XX CC anticonvulsant; antiarthritic; immunosuppressive; antihypertensive;
 XX CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 XX CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 XX CC antiviral; antifungal; antirheumatic; antihypertensive; and antianemic. The
 XX CC sequences can be used for determining the presence of or predisposition
 XX CC to, or preventing or treating pathological conditions associated with an
 XX CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 XX CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 XX CC used to treat cancers, proliferative disorders, neurodegenerative
 XX CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 XX CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 XX CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 XX CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 XX CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and

XX The invention relates to 2495 novel polynucleotides (I) and their encoded
CC polypeptides, sequences hybridizing to these nucleotides, sequences
CC encoding partial polypeptides and sequences having 70% or 90% identity to
CC the nucleotide and protein sequences. The nucleotides and polypeptides
CC are useful as diagnostic markers or therapeutic target for the diseases
CC or morbid states. They are also useful for treating osteoporosis,
CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC dementia and various cancers. This sequence corresponds to a protein
CC of the invention.
SQ Sequence 153 AA;
Query Match 24.9%; Score 806; DB 8; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.8e-46;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 462 MESSVAGERSVLAPESMRDHSVEPLRDPNPSDLLENLIDSVFSKRAKLEDEKRRKRD 521
DB 1 MESSVAGERSVLAPESMRDHSVEPLRDPNPSDLLENLIDSVFSKRAKLEDEKRRKRD 60
QY 522 IGRIRERILQRLQLMYKKKGIOESEPEVTSFPPEPDVESLMTTFPLPVVAFGRPLPK 581
DB 61 IGRIRERILQRLQLMYKKKGIOESEPEVTSFPPEPDVESLMTTFPLPVVAFGRPLPK 120
QY 582 LTFPQNELPMLDBRSRCLEIQKKQTPHRTCRK 614
DB 121 LTFPQNELPMLDBRSRCLEIQKKQTPHRTCRK 153
RESULT 9
AAU30711
ID AAU30711 standard; protein; 195 AA.
XX AAU30711;
XX 18-DEC-2001 (first entry)
XX Novel human secreted protein #1202.
XX Human; Vaccination; gene therapy; nutritional supplement;
XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX Homo sapiens.
XX WO200179449-A2.
XX 25-OCT-2001.
XX 16-APR-2001; 2001WO-US008656.
XX 18-APR-2000; 2000US-0052929.
XX 26-JAN-2001; 2001US-00770160.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-611725/70.
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy.
XX Claim 20; Page 334; 765pp; English.
XX The invention relates to novel human secreted polypeptides. The
XX polypeptides and antibodies to the polypeptides are useful for
XX determining the presence of or predisposition to a disease associated
XX with altered levels of polypeptide. The polypeptides are also useful for
XX identifying agents (agonists and antagonists) that bind to them. Cells
XX expressing the proteins are useful for identifying a therapeutic agent
XX for use in treatment of a pathology related to aberrant expression or

CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemia.
CC AAU29510-AAU33504 represent the amino acid sequences of novel human
CC secreted proteins of the invention
SQ Sequence 195 AA;
Query Match 11.7%; Score 377.5; DB 4; Length 195;
Best Local Similarity 46.3%; Pred. No. 1.9e-17;
Matches 94; Conservative 14; Mismatches 72; Indels 23; Gaps 8;
QY 10 AAAAPAGNPEORLDEYRAAALGSPDEDEGAABHPLPHRKLK-EPGPPLASSGGSPFA 68
DB 1 AAAAPAGNPEORLDEYRAAALGSPDEDEGAABHPLPHRKLK-EPGPPLASSGGSPFA 60
QY 69 PSPAGC-GGKGRGLLPAGAAP---GQDESWGGSVPLPCP--PPATKQAGIGG-----E 117
DB 61 PSPAGCGGKGGGLVTPGRGSPFAAGRELR---AVRCPCVPYRPPSPKXPALGGSIPQE 116
QY 118 PAAAGGCGSPREKYQAVLPITQGSVLAAPKEPFPAGDGGGAASPAATASDPAGPPPLPL 177
DB 117 PAAA-----PGSIRFVLPITQGSPPWRPKSLRPVLGTGVGRTPLPPLPPDPAGPPLPL 171
QY 178 PGPPPLAPTATAGTL--AASEGR 198
DB 172 PGPHSPRPPTGTPWRPARADGR 194
RESULT 10
ADFe0190
ID ADFe0190 standard; protein; 195 AA.
XX ADFe0190;
XX 12-FEB-2004 (first entry)
XX Human contig polypeptide sequence SEQ ID NO:2557.
XX biological activity; genetic engineering; hybridisation probe; oligomer;
XX primer; chromosome mapping; gene mapping; recombinant protein production;
XX human.
XX Homo sapiens.
XX WO2003080795-A2.
XX 02-OCT-2003.
XX 09-AUG-2002; 2002WO-US025485.
XX 09-AUG-2001; 2001US-0311261P.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Yang Y, Wang Z, Weng G, Ma Y;
XX WPI; 2003-876918/81.
XX N-PSDB; ADF59738.
XX New polynucleotides, useful as hybridization probes, oligomers or
XX primers, for chromosome or gene mapping, for the recombinant production
XX of proteins, and for generating antisense DNA or RNA.
XX Example 3; SEQ ID NO 2557; 571pp; English.
XX The present invention describes isolated polynucleotide sequences (I),
XX which encode polypeptides (II) with biological activity. Also described:

(1) a vector comprising (1); (2) an expression vector comprising (1); (3) a host cell genetically engineered to comprise (1) which is operatively associated with a regulatory sequence that modulates expression of (1) in the host cell; (4) a polypeptide (II) encoded by (1); (5) a composition comprising the polypeptide of (4) and a carrier; (6) an antibody directed against the polypeptide of (4); (7) detecting (1) or the polypeptide of (4) in a sample; (8) identifying a compound that binds to the polypeptide of (4); (9) producing the polypeptide of (4); and (10) a collection of polynucleotides comprising at least one of the polynucleotide sequences (1). The polynucleotides (1) can be used as hybridisation probes, oligomers or primers, for chromosome or gene mapping, for the recombinant production of proteins, and for generating antisense DNA or RNA. The present sequence represents a human contig polypeptide sequence, which is used in an example from the present invention.

SQ Sequence 195 AA;

Query Match 11.7%; Score 377.5; DB 7; Length 195;

Best Local Similarity 46.3%; Pred. No. 1.9e-17; Matches 94; Conservative 14; Mismatches 72; Indels 23; Gaps 8;

QY 10 AAAAPAGNPEORLDYERAAALGGDEPAGAAHFLPHRKUK-ERGPPLASQGGSPA 68
1 AAAAPAGNPEORLDYERAAALGGDEPAGAAHFLPHRKUK-ERGPPLASQGGSPA 60
QY PSPAGC-GGKGRGLL-PGAAAP---GOEESWGGSVLP-PCP--PPATKQAGIGG-----E 117
61 PSRAGCGGGGGLVTPGRGPRMAGREL---AVNCPYRVRPRPSKALGSLQPE 116
QY 118 PAAAGAGCSPRPYQAVLPITQSGSLVAAKEPTPWAGDKGAGSPAAATASDPAGPPLPL 177
117 PAA-----GPGSIRPVLPIQTGSPWRRPKSLRPVLGTRVGRTPPLPPDPAGPPLPL 171
QY 178 RGPPLAPRTAGTL--AASEGR 198
172 RGPSPRPPTGPRPARADGR 194
Db

RESULT 11

ADP84563 standard; protein; 60 AA.

XX ADP84563;

DT 09-SEP-2004 (first entry)

XX Human breast-specific protein #67.

DE Human breast-specific protein; breast cancer.

XX human; breast-specific protein; breast cancer.

OS Homo sapiens.

XX WO2004053077-A2.

PD 24-JUN-2004.

XX 05-DEC-2003; 2003WO-US038815.

XX 05-DEC-2002; 2002US-0431123P.

XX (DIAD-) DIADEXUS INC.

XX Macina RA, Turner LR, Sun Y, Chen H, Rodriguez M;

XX WPI; 2004-468848/44.

XX N-PSDB; ADP84445.

XX New breast specific nucleic acid molecules and polypeptides useful for

XX PT diagnosing, preventing or treating breast cancer, for producing

XX PT transgenic animals or cells, or for research purposes.

XX Claim 12; SEQ ID NO 162; 521pp; English.

CC The invention comprises the amino acid and coding sequences of human breast-specific proteins. The DNA and protein sequences of the invention are useful for the diagnosis, treatment and prevention of breast cancer. CC The present amino acid sequence represents a human breast-specific protein of the invention.

SQ Sequence 60 AA;

Query Match 8.2%; Score 264; DB 8; Length 60;

Best Local Similarity 94.2%; Pred. No. 2e-10; Matches 49; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 407 PSTHPEKAPSSIEDLPYSTETWYICRWHPSPPLPRSSPKKEETVA 458
1 PSTHPEKAPSSIEDLPYSTETWYICRWHPSPSPPLPRSSPKKEETVA 57
Db 6 PPAPEKAPSSIEDLPYSTETWYICRWHPSPSPPLPRSSPKKEETVA 57

RESULT 12

ABG75823 standard; protein; 1119 AA.

XX ABG75823;

DT 10-MAY-2003 (first entry)

XX Transporters and ion channels protein 5, TRICH-5.

KW Human; transporter and ion channel; TRICH; gene therapy;
KW cell proliferative disorder; transport disorder; neurological disorder;
KW muscle disorder; immunological disorder; amyotrophic lateral sclerosis;
KW cystic fibrosis; diabetes; Parkinson's disease; prostate cancer;
KW cardiac disorder; angina; Alzheimer's disease; amnesia; epilepsy;
KW schizophrenia; sickle cell anaemia; infertility; hyperglycaemia;
KW hypoglycaemia; hypercholesterolaemia; stroke; multiple sclerosis;
KW motor neuron disorder; prion disease; metabolic disease;
KW developmental disorder; central nervous system; cardiomyopathy;
KW hypertension; asthma; AIDS; allergy; anaemia; atherosclerosis;
KW atopic dermatitis; diabetes mellitus; osteoarthritis; osteoporosis;
KW rheumatoid arthritis; psoriasis; infection; trauma; hepatitis; cancer;
KW leukemia; lymphoma.

XX Homo sapiens.

XX WO2003016493-A2.

XX 27-FEB-2003.

XX 16-AUG-2002; 2002WO-US026323.

XX 17-AUG-2001; 2001US-0313242P.
XX 21-SEP-2001; 2001US-0324782P.
XX 02-OCT-2001; 2001US-0328184P.
XX 26-OCT-2001; 2001US-0345937P.
XX 01-NOV-2001; 2001US-035698P.
XX 13-NOV-2001; 2001US-0332804P.
XX 27-NOV-2001; 2001US-0333922P.
XX 26-APR-2002; 2002US-0375637P.
XX 03-MAY-2002; 2002US-0377444P.
XX 11-JUN-2002; 2002US-0388180P.

XX (INCY-) INCYTE GENOMICS INC.

XX Tang YT, Lal PG, Yue H, Baughn MR, Nguyen DB, Yao MG, Greene BD;
PI Borowsky ML, Lee S, Emerling BM, Xu Y, Beecha SD, Goryad AE;
PI Azimzai Y, Yue H, Elliott VS, Lee EA, Yang J, Lehr-Mason PM;
PI Ramkumar J, Lee SY, Paris M, Turner C, Furness M, Buchbinder JL;
PI Walia NK, Li JX, Forsythe JU, Griffin JA, Gietzen KU, Swarnakar A;
PI Hafalia AJA, Lindquist EA, Jiang X, Jackson AA, Wilson AD, Jin P;
PI Khare R, Marquis JP;

XX WPI; 2003-268319/26.

XX N-PSDB; ABX12008.

PT Novel human transporter and ion channel polypeptides and polynucleotides
PT for diagnosing, preventing or treating cell proliferative, transport,
XX neurological, muscle and immunological disorders.

PS Claim 1; Page 195-198; 253pp; English.

XX The invention discloses isolated polypeptides chosen from human
CC transporter and ion channel polypeptides, TRICH 1-26, a biologically
CC active or immunogenic fragment and the nucleic acids encoding them. Also
CC disclosed are isolated antibodies raised against the TRICH proteins,
CC methods for detecting a target polynucleotide in a sample and a
CC microarray where at least one element is a TRICH polynucleotide. The
CC proteins are useful for screening for agonists or antagonists, which can
CC then be used for treating a disease or condition associated with
CC decreased or overexpression of functional TRICH in a patient, for
CC screening for a compound that modulates the activity of the polypeptide
CC or that binds to the polypeptide or as an immunogen for preparing
CC antibodies. The polynucleotides are useful for screening for compounds
CC which alter expression of a target polynucleotide or for assessing
CC toxicity of a test compound. The polypeptides, polynucleotides,
CC modulators and antibodies are useful for diagnosis, treatment (e.g. gene
CC therapy) and prevention of cell proliferative, transport, neurological,
CC muscle and immunological disorders, such as amyotrophic lateral
CC sclerosis, cystic fibrosis, diabetes, Parkinson's disease, prostate
CC cancer, cardiac disorders, angina, Alzheimer's disease, amanesia,
CC epilepsy, schizophrenia, sickle cell anaemia, infertility,
CC hyperglycaemia, hypoglycaemia, hypercholesterolaemia, stroke, multiple
CC sclerosis, motor neuron disorder, prion disease, metabolic disease of the
CC nervous system, developmental disorders of the central nervous system,
CC cardiomyopathy, hypertension, asthma, AIDS, allergies, anaemia,
CC atherosclerosis, atopic dermatitis, diabetes mellitus, osteoarthritis,
CC osteoporosis, rheumatoid arthritis, psoriasis, infections, trauma,
CC hepatitis and cancers, including leukemia and lymphoma. The sequences
CC presented in ABG75819-ABG75844 are the TRICH proteins of the invention

XX Sequence 1119 AA;

XX Query Match 7.0%; Score 227.5; DB 6; Length 1119;
XX Best Local Similarity 23.2%; Pred. No. 2.2e-06;
XX Matches 117; Conservative 62; Mismatches 177; Indels 149; Gaps 22;

QY 19 PEORLYERAALAGPEDEFGAABAHFLPRHRLKKEGPRPLASSOGSS-----PAPBPAG 73
DB 448 PKREBELEQ-AQIKGVDPVGRBDGKEAPBEADQDRPGGCIAPVGAHNEHPVPDPKV 506
QY 74 CGSGKGRLLLP-----AGAPG-----QOESWGSVPLPCPPA 108
DB 507 VDEGDRVPEENKPRSRHAGKAPGVGOMAPRLPDSREKQEPGSGVGRPGQA 566
QY 109 TKQAG-----IGSPRAAGA-----GSPPRKYQAVLPQTGSLVAAAKS 148
DB 567 LBEAGDLPEBDPOKVPBEADGPAVOPAKEDLGPGRGLHPRP-QAVLSEQONGIAVGGSE 624
QY 149 PT---PWAGDKGGAAPAAATASDPAPGPRPLP-PCGPRPLAPRATAGTLAASGRKMSMR 203
DB 625 KAKGPPPPGAADDTQCPAEDSDHGKPRPLPAEKPAAGPGPLP-----BPPE 671
QY 204 KSPILGGSGSGASGSO-----AACTKQIILLDLLEQOQOOLQAKKEITELKSERDTL 257
DB 672 QRVVERAAGNGQAASQLEAGRAEMLDHVALVQV-KEGQVQO-----KRLDQGEKL 722
QY 258 LARIERMERMOVLKQDNEKERHKLFGQYTEERETSEKIKLECG-QPELSETSGTLP 316
DB 723 LAVIEQHKKEIQHQRDEDEDKPR-----QVEVHQEAGAVPQGEAPBEKARETVENLP 777
QY 317 PKFSGSGSGSKGHRKSPFSGTERKTPVKKLAPFSKVKTKTPRHS--PIKEPGCSLSE 374
DB 778 PLPLD-----PVLRAPG--RPAPSQUL-----NQRSSHSGVGRDPAG----- 816
QY 375 TVCKRELRSQETPEKPRSSVDTPPRLSTPOKGPSTHKEKAFSEIED-----LPVYSTT 429
DB 817 -----PDDGPPTEP--RAAQAKLRDGGQDAAPRAAGT 847

QY 430 EMYLCRMHQPSPSPPLPRESPPKE 454
DB 848 VKELPKGPQVPPDPAREAGPBE 872

RESULT 13

ID AAE20839 standard; protein; 720 AA.

AAE20839;

01-JUL-2002 (first entry)

Human gene 7 encoded secreted protein fragment, SEQ ID NO:101.

XX Human; secreted protein; gene therapy; human immunodeficiency virus; HIV;
XX immune disease; autoimmune disease; anaemia; multiple sclerosis; cancer;
XX rheumatoid arthritis; hyperproliferative disorder; melanoma; neoplasm;
XX Sezary syndrome; Gaucher's disease; neurological disease; cardiac arrest;
XX Alzheimer's disease; Parkinson's disease; Charcot-Marie-Tooth disease;
XX cardiovascular disorder; cerebrovascular disorder; tachycardia; angina;
XX thrombosis; ocular disorder; corneal infection; wound healing; cardiac;
XX vascular; thrombolytic; cyostatic; nootropic.

OS Homo sapiens.

PN WO200218435-A1.

PD 07-MAR-2002.

PF 17-JAN-2001; 2001MO-US001567.

PR 28-AUG-2000; 2000US-0228084P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Komatsuoli GA, Baker KP, Birse CE, Soppet DR;
PI Olsen HS, Moore PA, Wei P, Edner R, Duan DR, Shi Y, Choi GH;
PI Fliscella M, Ni J;

DR WPI; 2002-281060/32.

PT Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating e.g. Alzheimer's disease, cardio-
PT /cerebrovascular disorders and multiple sclerosis.

PS Disclosure; Page 18-19; 504pp; English.

CC AAD33237-AAD33280 represent cDNAs corresponding to 18 human secreted
CC protein genes, and AAE20793-AAE20836 represent the proteins they encode.
CC AAE20837-AAE20847 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 18 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of immune
CC or autoimmune diseases (e.g. HIV (human immunodeficiency virus)
CC infections, anemia, rheumatoid arthritis and multiple sclerosis),
CC cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
CC diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-
CC Marie-Tooth disease), cardiovascular or cerebrovascular disorders (e.g.
CC cardiac arrest, tachycardia, angina and thrombosis), infections caused by
CC bacteria, viruses and fungi and ocular disorders (e.g. corneal
CC infections). Secreted proteins of the invention can also be used to
CC promote wound healing, maintain organs before transplantation, support
CC cell culture of primary tissues, modulate differentiation of embryonic
CC stem cells, induce mesodermal tissue to differentiate in embryos,
CC modulate mammalian characteristics (e.g. height and weight), modulate the
CC catabolism, anabolism, energy storage, mental state, biorhythms, cardiac
CC rhythms, reproductive potential, hormonal levels appetite, memory and


```
QY 258 LARIERMERMOLVKDKNEKRNKLFQGYTEREBTELSEKIKLAC-QPELSETQTLR 316
DB 371 LAVIEQKHKEIHQORDEDEKPR-----QVEYHQBGAAVPRGQAEPEKARETVENLP 425
QY 317 PKPFGSGRSGKGHKRSPGSTERKTPVKKLADEFKVKTKTKHS--PIKEPPCSLSR 374
DB 426 PLPLD-----PVLAAPGQ---RPASQDL-----NQRSLHEBGPGRDPAG---- 464
QY 375 TVCKRELRSQETPEKPRSSVDTPRPLSTPQKGPSTHPKE--KAFSSEIDLPYLSTTEM 431
DB 465 -----PDGSGDTSPRAAGKLRDQGXDAAPRAAGTVK 497
QY 432 YLCRMHQPPSPPLPLESSPKKE 454
DB 498 ELPKGPEQVVPDPAREAGPPEE 520

RESULT 15
AAE20799
ID AAE20799 standard; protein; 973 AA.
XX
AC AAE20799;
XX
DT 01-JUL-2002 (first entry)
XX
DE Human gene 7 encoded secreted protein HMVP38, SEQ ID NO:61.
XX
KW Human; secreted protein; gene therapy; human immunodeficiency virus; HIV;
KW immune disease; autoimmune disease; anaemia; multiple sclerosis; cancer;
KW rheumatoid arthritis; hyperproliferative disorder; melanoma; neoplasm;
KW Sezary syndrome; Gaucher's disease; neurological disease; cardiac arrest;
KW Alzheimer's disease; Parkinson's disease; Charcot-Marie-Tooth disease;
KW cardiovascular disorder; cerebrovascular disorder; tachycardia; angina;
KW thrombosis; ocular disorder; corneal infection; wound healing; cardiac;
KW vascular; thrombolytic; cytostatic; nootropic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..34
FT /label=Signal_peptide
FT Protein 35..973
FT /label=Mature_secreted_protein
PN WO200218435-A1.
XX
PD 07-MAR-2002.
XX
PP 17-JAN-2001, 2001WO-US001567.
XX
PR 28-AUG-2000; 2000US-0228084P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Komatsuji GA, Baker KP, Birse CE, Soppet DR,
PI Olsen HS, Moore PA, Wei P, Edner R, Duan DR, Shi Y, Choi GH,
PI Fiscella M, Ni J;
XX
DR WPI: 2002-281060/32.
DR N-PSDB; AAD33243.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating e.g. Alzheimer's disease, cardio-
PT /cerebrovascular disorders and multiple sclerosis.
XX
PS Claim 11; Page 451-454; 504pp; English.
XX
CC AAD33237-AAD33280 represent cDNAs corresponding to 18 human secreted
CC protein genes, and AAE20793-AAE20836 represent the proteins they encode.
CC AAE20837-AAE20847 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
```

```
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 18 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of immune
CC or autoimmune diseases (e.g. HIV (human immunodeficiency virus)
CC infections, anaemia, rheumatoid arthritis and multiple sclerosis),
CC cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
CC diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-
CC Marie-Tooth disease), cardiovascular or cerebrovascular disorders (e.g.
CC cardiac arrest, tachycardia, angina and thrombosis), infections caused by
CC bacteria, viruses and fungi and ocular disorders (e.g. corneal
CC infections). Secreted proteins of the invention can also be used to
CC promote wound healing, maintain organs before transplantation, support
CC cell culture of primary tissues, modulate differentiation of embryonic
CC stem cells, induce mesodermal tissue to differentiate in embryos,
CC modulate mammalian characteristics (e.g. height and weight), modulate the
CC catabolism, anabolism, energy storage, mental state, biorhythms, cardiac
CC rhythms, reproductive potential, hormonal levels appetite, memory and
CC stress. They can also be used as an additive to increase or decrease
CC storage capabilities and nutritional content of food. The present
CC sequence represents a human secreted protein of the invention
XX
SQ Sequence 973 AA:
Query Match 7.0%; Score 226.5; DB 5; Length 973;
Best Local Similarity 23.1%; Pred. No. 2,1e-06;
Matches 116; Conservative 61; Mismatches 181; Indels 145; Gaps 21;
QY 19 PEORLYERAAALGSPDEPGAAEAHFLPRHRLKRGPRPLASSOGS-----PAPSPAG 73
DB 303 PKREBLEQ-AQIKGVVDVDFREDGKABEADLRDGGIAYVGEAHNHEPVPVHDKY 361
QY 74 CCGKGRGLLP-----AGAPG-----QDEBSWGSVPLPCPPPA 108
DB 362 VDEGDQDRVPEBENKRPNSHAGKAPGVQGMARPLDSEBKEQDEPGEVGRPQQAQ 421
QY 109 TKQAG-----TGEPPAAGA-----GCSPPKYQAVLPITQGSVLAANE 148
DB 422 LBEAGDLPEDPQKVPADQPAVQPAKEDLGPDRGLHPRP-QAVLSEQNGLAVGGGE 479
QY 149 PT---PWAGDKGAAPATAAPDAPGRPLPL--PGPPPLAPATAGTLAASGRKMSNR 203
DB 480 KAKGPPPGAAADTQAPABSDHSGKPPPLPAKPRAGELP-----EPRE 526
QY 204 KSPFGGSGSGASGQ-----AACLKQILLDLTEOQOQOLQAKEKEIEBKSRDYL 257
DB 527 QRDVERAGNGQAASQLEBAGRAMLDHVLQV-IEQVQVQ-----KRLDQQEK 577
QY 258 LARIERMERMOLVKDKNEKRNKLFQGYTEREBTELSEKIKLAC-QPELSETQTLR 316
DB 578 LAVIEQKHKEIHQORDEDEKPR-----QVEYHQBGAAVPRGQAEPEKARETVENLP 632
QY 317 PKPFGSGRSGKGHKRSPGSTERKTPVKKLADEFKVKTKTKHS--PIKEPPCSLSR 374
DB 633 PLPLD-----PVLAAPGQ---RPASQDL-----NQRSLHEBGPGRDPAG---- 671
QY 375 TVCKRELRSQETPEKPRSSVDTPRPLSTPQKGPSTHPKE--KAFSSEIDLPYLSTTEM 431
DB 672 -----PDGSGDTSPRAAGKLRDQGXDAAPRAAGTVK 704
QY 432 YLCRMHQPPSPPLPLESSPKKE 454
DB 705 ELPKGPEQVVPDPAREAGPPEE 727
```

Search completed: March 23, 2005, 15:28:35
Job time : 171 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2005, 15:14:35 ; Search time 180 Seconds
(without alignments)
1746.759 Million cell updates/sec

Title: US-10-054-935-2

Sequence: 1 MTMR5AVFKAAAPAGANPE.....RSRCRLIEOKKQTPHRTCRK 614

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3115	96.2	616	2	Q6PDM1 mus musculu
2	3110	96.0	616	2	Q9CXF9 mus musculu
3	2913	90.0	598	2	Q8BPN3 mus musculu
4	2306	71.2	463	2	Q80XSO mus musculu
5	2155	66.6	413	2	Q68DK7 mus musculu
6	1711	52.8	370	2	Q9D5C9 mus musculu
7	669	20.7	125	2	Q69Z03 mus musculu
8	257	7.9	274	2	Q7Q1Y2 anopheles g
9	222	6.9	919	2	Q6NZL0 mus musculu
10	215	6.6	1126	2	Q7PXZ6 anopheles g
11	213	6.6	1039	1	MSL1_DROME
12	212.5	6.6	1130	2	Q81ZL8 mus musculu
13	209.5	6.5	743	2	Q9BUD0 mus musculu
14	209.5	6.5	1048	2	Q96FT1 mus musculu
15	209.5	6.5	1130	2	Q6NTE6 mus musculu
16	206.5	6.4	802	2	Q8K3Z7 mus musculu
17	203.5	6.3	2365	2	Q8NHDO mus musculu
18	203.5	6.3	2429	2	Q8NHCO mus musculu
19	203.5	6.3	2429	2	Q8NHCO mus musculu
20	203.5	6.3	2432	2	Q8TDE9 mus musculu
21	202	6.2	675	2	Q7YUR6 mus musculu
22	202	6.2	1284	2	Q154S0 mus musculu
23	200.5	6.2	1021	2	Q154S1 mus musculu
24	200	6.2	1026	1	NFH_HUMAN
25	200	6.2	3937	1	BSN_RAT
26	199	6.1	1200	2	Q7SBD4 mus musculu
27	197.5	6.1	831	1	NFH_RAT
28	197.5	6.1	1072	2	Q354E2 mus musculu
29	197.5	6.1	2157	2	Q9Z1R1 mus musculu
30	197.5	6.1	2158	2	Q7TSC1 mus musculu
31	197.5	6.1	3941	1	BSN_MOUSE

32	TAF4	197	6.1	1083	1	Q6GQX2	HUMAN	000268	homo sapien
33		197	6.1	1323	2	Q6GQX2	mus musculu	Q6GQX2	mus musculu
34		196.5	6.1	812	2	Q6P181		Q6P181	homo sapien
35		196.5	6.1	821	2	Q96JH3	homo sapien	Q96JH3	homo sapien
36		196.5	6.1	825	2	Q7RZ46		Q7RZ46	neutrospora
37		196.5	6.1	2715	1	ML14	HUMAN	Q9unm6	homo sapien
38		196	6.1	977	2	Q68A20		Q68A20	homo sapien
39		195.5	6.0	804	2	Q6ZP11		Q6ZP11	mus musculu
40		195.5	6.0	858	2	Q6ZP27		Q6ZP27	mus musculu
41		194.5	6.0	740	2	Q8NC88		Q8nc88	homo sapien
42		194.5	6.0	1123	2	Q9DAD5		Q9dad5	mus musculu
43		194	6.0	1479	2	Q8ZFP7		Q8zfp7	streptomyce
44		193.5	6.0	520	2	Q61078		Q61078	mus musculu
45		193.5	6.0	1197	1	SNIP_RAT		Q9qxy2	rattus norv

ALIGNMENTS

RESULT 1	Q6PDM1	PRELIMINARY	PRT;	616 AA.
ID	Q6PDM1	05-JUL-2004 (TREMBlrel. 27, Created)		
AC	Q6PDM1	05-JUL-2004 (TREMBlrel. 27, Last sequence update)		
DT	05-JUL-2004 (TREMBlrel. 27, Last sequence update)			
DT	05-JUL-2004 (TREMBlrel. 27, Last sequence update)			
DE	RIKEN CDNA 4121402D02.			
GN	Name=4121402D02.Rik;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6; TISSUE=Mouse;			
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;			
RA	St/ausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.F., Collins F.S., Wagner C.M., Schuler G.D.,			
RA	Mitschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,			
RA	Brownstein M.J., Udell T.B., Tothiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,			
RA	Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,			
RA	Villalón D.K., Muzny D.W., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,			
RA	Krzywnski M.I., Skalska U., Smailus D.E., Schenck A., Schein J.E.,			
RA	Jones S.J., Marra M.A.,			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6; TISSUE=Mouse;			
RA	Strausberg R.,			
RL	Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC058629; AAHS8629.1; -			
SO	SEQUENCE 616 AA; 67319 MW; 2679AF230DBC0A7D CRC64;			
QY	Query Match	96.2%; Score 3115; DB 2; Length 616;		
QY	Best Local Similarity	96.1%; Pred. No. 6; ee-120;		
QY	Matches 592; Conservative 10; Mismatches 12; Indels 2; Gaps 2;			
DB	1 MTMR5AVFKAAAPAGANPEORLDYERAAALGPEDEPGAAEAHFLPRHKKTERGPPIA 60			
QY	1 MTMR5AVFKAAAPAGANPEORLDYERAAALGPEDESGAAEAHFLPRHKKTERGPPIA 60			
QY	61 SSQGGSPAPAPAGC-GGKGRGLLPAGAAPGQGEESGGSVPLPCPPATKAGIGGEPA 119			

```
Db 61 SSQGGSPSPSPACGGCGKRGILLPAGAAPGQOEBSWGSVPLPCPPPATKQIGGEPV 120
Qy 120 AAGAGCSPRPKYQAVLPLOTGSL-VAAAKPTPWADKGGAASPAATADPAGPPLPLP 178
Db 121 AAGAGCSPRPKYQAVLPLOTGSLVVAAAKPTPWADKGGAAPPAATADPAGPPLPLP 180
Qy 179 GPPPLAPPTAGTTLAASBGRWKSIRKSPLOGGGGSGASSQAACLKQILLQDLIRQOOO 238
Db 181 GPPPLAPPTAGTTLAASBGRWKSIRKSPLOGGGGSGASSQAACLKQILLQDLIRQOOO 240
Qy 239 QLOAKEKEIEBELKSEBDTLIARIERMERRMQLVKDNEKERHKLQGYETEEBETELSE 298
Db 241 QLOAKEKEIEBELKSEBDTLIARIERMERRMQLVKDNEKERHKLQGYETEEBETELSE 300
Qy 299 KILLECOPELSETQTLPPKPFSCGSGKHKRKSPPGSTERKTTPVKCLAPESKYTKT 358
Db 301 KILLECOPELSETQTLPPKPFSCGSGKHKRKSPPGSTERKTTPVKCLAPESKYTKT 360
Qy 359 PKHSPITKEPCGSLSETVCKRELRSQETPKPRSSVDTPLRLSTPOKGPSTHKEKAFSS 418
Db 361 PKHSPITKEPCGSLSETVCKRELRSQETPKPRSSVDTPLRLSTPOKGPSTHKEKAFSS 420
Qy 419 EIEDLPYLSTTEWYLGRMHQPPSPPLPLRSSPKCEETVARCLMPSVAGETSVLAVPSW 478
Db 421 EMBDLPYLSTTEWYLGRMHQPPSPPLPLRSSPKCEETVARCLMPSVAGETSVLAVPSW 480
Qy 479 RDHSVEPLADPNBDDLENDLSVFSKRAKLELDEKRRKRWDIQRIREQRIILQRLQLM 538
Db 481 RDHSVEPLADPNBDDLENDLSVFSKRAKLELDEKRRKRWDIQRIREQRIILQRLQLM 540
Qy 539 YKKKGQIESPEPTSPFPEDDVESLMTPLFVVAAGRPPLKTLPTONFELPWLDERSRC 598
Db 541 YKKKGQIESPEPTSPFPEDDVESLMTPLFVVAAGRPPLKTLPTONFELPWLDERSRC 600
Qy 599 RLEIOKQOTPHRTCRK 614
Db 601 RLEIOKQOTPHRTCRK 616

RESULT 2
Q9CXF9 PRELIMINARY; PRT; 616 AA.
ID 09CXF9
AC 09CXF9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
DE library, clone:4121402D02 product:hypothetical protein, full insert
DE sequence.
GN Name=4121402D02Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=92729253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN PANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA The PANTOM Consortium;
```

```
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Saeki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:11757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Atakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Futuno M.,
RA Hanaoka T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Nunazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL:AK014463; BMB29369.1;
DR MGI:MGI:1921276; 4121402D02Rik.
KW Hypothetical protein.
SQ SEQUENCE 616 AA; 67332 MW; 221C84C306C7D07D CRC64;

Query Match 96.0%; Score 3110; DB 2; Length 616;
Best Local Similarity 95.9%; Pred. No. 1,1e-119;
Matches 591; Conservative 10; Mismatches 13; Indels 2; Gaps 2;
```

```

Db 301 KILEROPELCEISOALPSKPFSCGSGKHAKRTPEGNTERRKNPVKKLAPERSKVTKT 360
Qy 359 PKHSPKIEBPGSLSETVCKRELRSOETPEKPRSSVDTPLPRLSTPOKGPSTHPREKAFSS 418
Db 361 PKHSPKIEBPGSLSETVCKRELRSOETPEKPRSSVDTPLPRLSTPOKGPSTHPREKAFSS 420
Qy 419 EIEDLPLSTTMYLTCRMHOPPPSPPLPLRSSPKKEETVARCLMPSSVAGETSVLAVPSW 478
Db 421 EMBEDLPLSTTMYLTCRMHOPPPSPPLPLRSSPKKEETVARCLMPSSVAGETSVLAVPSW 480
Qy 479 RDSHVEPLRPNPNDLLENLDDSVFSGRHAKELEDEKRRKRWDTQRIREQRIIQRQLRM 538
Db 481 RDSHVEPLRPNPNDLLENLDDSVFSGRHAKELEDEKRRKRWDTQRIREQRIIQRQLRM 540
Qy 539 YKKKGIOESEPEVTSPFPPEPDVESLMTPLPLPVAFGRPLPKLTPONFELPWLDESRRC 598
Db 541 YKKKGIOESEPEVTSPFPPEPDVESLMTPLPLPVAFGRPLPKLTPONFELPWLDESRRC 600
Qy 599 RLEIQKKQTPHRTCRK 614
Db 601 RLEIQKKQTPHRTCRK 616

RESULT 3
Q8BPN3 PRELIMINARY; PRT; 598 AA.
ID Q8BPN3;
AC 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
DE library, clone: E130302B09 product: hypothetical protein, full insert
DE sequence.
GN Name=E121402D02Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RA The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;

```

```

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama T., Nishino T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujimake S., Inoue K., Togawa Y., Izawa M., Ohara E., Matshiki M.,
RA Yoneda Y., Ishikawa K., Ozawa K., Tanaka T., Matsunura S., Kawai U.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RA Adachi U., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirokane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katon H., Kawai U., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akanita S., Takeda Y., Tanaka T.,
RA Tomatsu A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK053719; BAC35489.1; -
DR MGD; MGI:1921276; 4121402D02Rik.
KW Hypothetical protein.
SQ
Sequence 598 AA; 65100 MW; 9F10EEBA9AA7ACCS CRC64;

Query Match 90.0%; Score 2913; DB 2; Length 598;
Best Local Similarity 93.3%; Pred. No. 1,2e-11;
Matches 556; Conservative 11; Mismatches 11; Indels 18; Gaps 3;

Qy 1 MTRSAVFKAAPADAGNPPQRLDYERAAALGPEDEPGAABHFLPRHKLKEGPPLA 60
Db 1 MTRSAVFKAAPADAGNPPQRLDYERAAALGPEDEPGAABHFLPRHKLKEGPPLA 60
Qy 61 SSQGGSPAPBAGC-GSGKRGILLPAGAPGQOESWGSGVPLPCPPATKOGTGEPA 119
Db 61 SSQGGSPSPBAGCGGKRGILLPAGAPGQOESWGSGVPLPCPPATKOGTGEPA 120
Qy 120 AAGAGCSPPRYOAVLPIQTGSL-VAAKEPTWAGDKGGAAPATASDPAGPPPLP 178
Db 121 AAGAGCSPPRYOAVLPIQTGSL-VAAKEPTWAGDKGGAAPATASDPAGPPPLP 180
Qy 179 GPPPLPATATAGTLAASGGRKMRKSPICGGGSGGASQAACIKOILLQLDLIEQOOQ 238
Db 181 GPPPLPATATAGTLAASGGRKMRKSPICGGGSGGASQAACIKOILLQLDLIEQOOQ 240
Qy 239 QLOAKKEIEBELSEBDTLARIERMERBMOVKKNEKERHLLFOGYETEEETLSE 298
Db 241 QLOAKKEIEBELSEBDTLARIERMERBMOVKKNEKERHLLFOGYETEEETLSE 300
Qy 299 KILEROPELSETSQTLPPKPFSCGSGKHAKRSPPGSTERKTPVKKLAPERSKVTKT 358
Db 301 KILEROPELCEISOALPSKPFSCGSGKHAKRTPEGNTERRKNPVKKLAPERSKVTKT 360
Qy 359 PKHSPKIEBPGSLSETVCKRELRSOETPEKPRSSVDTPLPRLSTPOKGPSTHPREKAFSS 418
Db 361 PKHSPKIEBPGSLSETVCKRELRSOETPEKPRSSVDTPLPRLSTPOKGPSTHPREKAFSS 420
Qy 419 EIEDLPLSTTMYLTCRMHOPPPSPPLPLRSSPKKEETVARCLMPSSVAGETSVLAVPSW 478
Db 421 EMBEDLPLSTTMYLTCRMHOPPPSPPLPLRSSPKKEETVARCLMPSSVAGETSVLAVPSW 480
Qy 479 RDSHVEPLRPNPNDLLENLDDSVFSGRHAKELEDEKRRKRWDTQRIREQRIIQRQLRM 538
Db 481 RDSHVEPLRPNPNDLLENLDDSVFSGRHAKELEDEKRRKRWDTQRIREQRIIQRQLRM 540
Qy 539 YKKKGIOESEPEVTSPFPPEPDVESLMTPLPLPVAFGRPLPKLTPONFELPWLDE 594
Db 541 YKKKGIOESEPEVTSPFPPEPDVESLMTPLPLPVAFGRPLPKLTPONFELPWLDE 580

```

```
RESULT 4
Q08XS0 PRELIMINARY; PRT; 463 AA.
ID Q08XS0
AC Q08XS0;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE 4121402D02Rik protein.
GN Name=4121402D02Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=1090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.U., Ustin T.B., Toshiyuki S., Abramson R.D., Mullany S.J.,
RA Rana S.S., Loguellano N.A., Peters G.D., Dickson M.C.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Goughwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043039; AAH43039.1; -.
DR EMBL; BC055715; AAH55715.1; -.
DR MGI; MGI:1921276; 4121402D02Rik.
SQ SEQUENCE 463 AA; 49206 MW; C52E141065B00CF2 CRC64;

Query Match 71.2%; Score 2306; DB 2; Length 463;
Best Local Similarity 95.7%; Pred. No. 5.7e-87;
Matches 440; Conservative 8; Mismatches 10; Indels 2; Gaps 2;

QY 1 MTRSAVFKAAPAGANPEORLDYERAAALGSPEDPEGAABAFPLPRHKLKEPGLA 60
DB 1 MTRSAVFKAAPAGANPEORLDYERAAALGSPEDPEGAABAFPLPRHKLKEPGLA 60
QY 61 SSQGGSPAPSPAGC-GGKGGILLPAGAAPGQGEESWGSPPLPCPPATKAGIGEP 119
DB 61 SSQGGSPAPSPAGC-GGKGGILLPAGAAPGQGEESWGSPPLPCPPATKAGIGEP 120
QY 120 AAGAGCSPPPKYQAVPIQTGSL-VAAAKEPTWAGDKGGAAPATASPPAPPLPLP 178
DB 120 AAGAGCSPPPKYQAVPIQTGSL-VAAAKEPTWAGDKGGAAPATASPPAPPLPLP 180
QY 179 GPPPLAFTYAGTAASEGKSMRKSPLGGGGSGASQAACLKQILLLQDLIRGQQQ 238
DB 179 GPPPLAFTYAGTAASEGKSMRKSPLGGGGSGASQAACLKQILLLQDLIRGQQQ 240
```

```
RESULT 5
Q68DK7 PRELIMINARY; PRT; 413 AA.
ID Q68DK7
AC Q68DK7;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Hypothetical protein DKFZp686J17211.
GN Name=DKFZp686J17211;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Rectum tumor;
RG The German CDNA Consortium;
RA Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Robo G., Han M., Wiemann S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR749360; CAH18213.1; -.
SQ SEQUENCE 413 AA; 47603 MW; 29BA8726890A80BE CRC64;

Query Match 66.6%; Score 2155; DB 2; Length 413;
Best Local Similarity 99.8%; Pred. No. 7.5e-81;
Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 202 MKSPVGGGGSGASSQAACLKQILLLQDLIRGQQQQAQAKEKEIEBKSERDTLLARI 261
DB 1 MKSPVGGGGSGASSQAACLKQILLLQDLIRGQQQQAQAKEKEIEBKSERDTLLARI 60
QY 262 ERMERMMQVKKDNEKERHKLFGYETBEREETELSEKIKLBCQPLSETSQTLPPKPS 321
DB 61 ERMERMMQVKKDNEKERHKLFGYETBEREETELSEKIKLBCQPLSETSQTLPPKPS 120
QY 322 CGRSKGHRKSPFGSTERTKTPVKLAPEFSKYKTTTPHSPKESPLSTVCKREL 381
DB 121 CGRSKGHRKSPFGSTERTKTPVKLAPEFSKYKTTTPHSPKESPLSTVCKREL 180
QY 382 RSQETEKRRSSVDTPRLSTPQKGPSTHPKEKAFSEIEIDLPLYLSTTEMVLCRMHQPP 441
DB 181 RSQETEKRRSSVDTPRLSTPQKGPSTHPKEKAFSEIEIDLPLYLSTTEMVLCRMHQPP 240
QY 442 SPLPLESSPKKEETVARCLMPSSVAGTSSVLAVPWRDHSVLEPLDPNPDLLENLDD 501
DB 241 SPLPLESSPKKEETVARCLMPSSVAGTSSVLAVPWRDHSVLEPLDPNPDLLENLDD 300
QY 502 VRSKRAKLEDKRRKRDQIRQRIQLRLQRLMYKKKGIQSEBPEVTSFFPPDDV 561
DB 301 VRSKRAKLEDKRRKRDQIRQRIQLRLQRLMYKKKGIQSEBPEVTSFFPPDDV 360
QY 562 ESMITPPLPVYAFGRPLKLPONFELPWLBERSKRLIEIOKKOPPHRTCK 614
DB 361 ESMITPPLPVYAFGRPLKLPONFELPWLBERSKRLIEIOKKOPPHRTCK 413
```


RESULT 6
ID 09D5C9 PRELIMINARY; PRT; 370 AA.
AC 09D5C9;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930463F05 product:hypothetical protein, full insert sequence.
GN Name=4121402D02Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P., Kono H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kaishwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Katsuhiki M., Itoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H., Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M., Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoaka T., Hori F., Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H., Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numasaki R., Ono M., Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,

RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK015496; BMB29868.1; -
DR MGD: MG11921276; 4121402D02Rik.
KW Hypothetical protein.
SQ SEQUENCE 370 AA; 43184 MW; 2347FA88F9B5654 CRC64;
Query Match 52.8%; Score 1711; DB 2; Length 370;
Best Local Similarity 89.3%; Pred. No. 9.2e-63;
Matches 324; Conservative 10; Mismatches 13; Indels 16; Gaps 1;
QY 252 SERDTLLARIERMERMOVLVKDKNEKRHLFGYETETETETELSEKILAECPLELSET 311
DB 24 TNNKLLARIERMERMOVLVKDKNEKRHLFGYETETETETELSEKILAECPLELSET 83
QY 312 SQTLPKPPSCGSGGKHKRKSPPGSTERTKTPVKKLAPFSSKYKTKTPKISPIKEPCGS 371
DB 84 SQALPKPPSCGSGGKHKRKSPPGSTERTKTPVKKLAPFSSKYKTKTPKISPIKEPCGS 143
QY 372 LSETVCKRELRSGETEKPPSSVDTTPRLSTPKGSPTHPKAFSEIEDLPYLSTTEM 431
DB 144 ISETVCKRELRSGETEKPPSSVDTTPRLSTPKGSPTHPKAFSEIEDLPYLSTTEM 203
QY 432 YLCRMHOPPPSPPLRESSPKKEETVARCLMPSSVAGETSVLAVPSWRDHSVEPLRDPNP 491
DB 204 YLCRMHOPPPSPPLRESSPKKEETV-----ALPSRDSVVEPLRDPNP 247
QY 492 SDILENLDSSVFSKRAKLELDKRRRRWDIQRIRQRILQRLQLMRYKKKGIOESPEV 551
DB 248 SDILENLDSSVFSKRAKLELDKRRRRWDIQRIRQRILQRLQLMRYKKKGIOESPEV 307
QY 552 TSFPPEDDVESLMTPLPVAFGRLPLKLTQNPFLPLDLSRRRLKIQKQTHRT 611
DB 308 TSFPPEDDVESLMTPLPVAFGRLPLKLTQNPFLPLDLSRRRLKIQKQTHRT 367
QY 612 CRK 614
DB 368 CRK 370
RESULT 7
ID 069203 PRELIMINARY; PRT; 125 AA.
AC 069203;
DT 25-OCT-2004 (TEMBLrel. 28, Created)
DT 25-OCT-2004 (TEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TEMBLrel. 28, Last annotation update)
DE Hypothetical protein DKFZp568B1922 (Fragment).
GN Name=DKFZp568B1922.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RG The German cDNA Consortium;
RA Koehler K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A., Fobio G., Han M., Wiemann S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL049450; CH10734.1; -
KW Hypothetical protein.
FT NON TER 1
FT NON TER 125
SQ SEQUENCE 125 AA; 14065 MW; C68F65B36FE59E03 CRC64;
Query Match 20.7%; Score 669; DB 2; Length 125;
Best Local Similarity 100.0%; Pred. No. 1.2e-20;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 334 PFGSTERKTPVKKLAPFSSKYKTKTPKHSPIKEPCGSLETVCKRELRSGETEKPPSS 393

```
Db 1 PGGSTRKTRKPVKLAEPFESKVKTKTKRKHPIKEPFGCSLSEITVCKELINSQETPEKPRSS 60
Qy 394 VDPPELSTPQKGPSTHPKEKASSEIEDLPYLSTTEMYICRMHQPPSPPLPRLRESSPKK 453
Db 61 VDPPELSTPQKGPSTHPKEKASSEIEDLPYLSTTEMYICRMHQPPSPPLPRLRESSPKK 120
Qy 454 EETVA 458
Db 121 EETVA 125

RESULT 8
Q701Y2 PRELIMINARY; PRT; 274 AA.
ID 0701Y2;
AC 0701Y2;
DT 01-MAR-2004 (TEMBLrel. 26, Created)
DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)
DE AGC88168.
GN Name=agc56300; ORFNames=ENSNAGC00000018620;
OS Anopheles gambiae str. PST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RL Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008980; EAA14384.1; -.
SQ SEQUENCE 274 AA; 30127 MW; BA21535A1AD07464 CRC64;

Query Match 7.9%; Score 257; DB 2; Length 274;
Best Local Similarity 30.8%; Pred. No. 0.0011;
Matches 72; Conservative 41; Mismatches 79; Indels 42; Gaps 8;

Qy 362 SPIKEPSCGSLSE-TVCKRELRSQETPEKPRSSVDPPELSTPQKGPSTHPKEKASSEI 420
Db 6 NPTLHPSGNSGSSGAAR-----RNCKPTAAVYVTKQYVSSGKMDATYAEI 54
Qy 421 EDLPYLSTTEMYICRMHQPPSPPLPRLRESSPKKEETVARCLMPSSVAGETSVLAIVSWRD 480
Db 55 EKLISNEAAELRIPSW-----TVIEDDDDDPGCGDPPGCGEPEPSVGA----- 98
Qy 481 HAYEPRLDRPSPDLLNLDSVFSKHAKELEDEKRRKWDIORIRQRLQLRLMYK 540
Db 99 --SEPSR-----ENISDEAYAKRHTKLEIDERRRKKWDVORIREQKHIERLKRRQLK 148
Qy 541 KKGII-GESEPEVTSFPPEPDVSLMTTPPLPVAFGRPLPKLTPON--PELPM 591
Db 149 EQVEBEAQKAITTYFTYDTLKYVLVDVPAQAGELIP-LLPYNGGFSLPM 201

RESULT 9
Q6NZLO PRELIMINARY; PRT; 919 AA.
ID 06NZLO;
AC 06NZLO;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TEMBLrel. 27, Last annotation update)
DE RIKEN CDNA 6330407J23.
GN Name=6330407J23Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
```

```
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheef C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson C.M., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC066079; AAH66079.1; -.
SQ SEQUENCE 919 AA; 100506 MW; C0274B0BCED1FCBE CRC64;

Query Match 6.9%; Score 222; DB 2; Length 919;
Best Local Similarity 22.3%; Pred. No. 0.14;
Matches 147; Conservative 79; Mismatches 228; Indels 204; Gaps 29;

Qy 12 AAPAGNPPQRLDYERAAAL-----GGEDEPGAAHFLPRHKKLKEGPPLASQ 63
Db 113 AVPGAIVPAPAGAGSAAITLAVSFPBSQGPPEAP-----REL-ESGP-----SK 156
Qy 64 GGSPPAP-SPAGCGGCKRGILLPAGAAPGQESWGSGVLPCCPP----- 106
Db 157 VGEPPPLGGVGGGEGEGG---AGGPPGDRE---GGAQPPPPPRGKGVATRGSS 208
Qy 107 -----PATKQAGIGGEPAAGA-----GCSPPRYQVAVLPIDTG 140
Db 209 VAEGVSPSPPTAATSKTPPBGSRNBSGSGTSGSGGSGSYWKEGCLQSELIDPHLKKEA 268
Qy 141 SLVAAAKEPTPAGDGKGAASPAATSDPAQPPPLPGLPPLAPATAGTAAASGRWK 200
Db 269 AAAAAAQAQHTKNGGSGSSPVAQA--PAICEPLVVPSSPMA-----AAAGPQG 318
Qy 201 SMKRSPLGGGGSGASSQAAC--LKOILLQLDLIQOQOOLQAKKEITBELKSERDTLL 258
Db 319 SAE-----GNSGGGMAQAAPSSQPHSQLOQEOEDMOEEMEKLEENETLKNEIDELR 372
Qy 259 ARIERY-----ERRQVYKXNKEK-----HKLFQGVETEE----- 290
Db 373 TEMDEMRDTFFBEDACQLQEMRHEIRANKNCRIIOLYLRKARRKRLRAYQTGEIDGELL 432
Qy 291 ---REETELSEKIKLECOPELSETSQLPKPFGSCGSGKHKRKSPPFSTER----- 340
Db 433 RSLIEDQLKAKOVSAVLIHLELNVBE-----KRTTDEENETLQQLTE 476
Qy 341 -KTPVKKLAPEPSKYTKTPKHSPIKEPCGSLSEITVCKRELRSQETPEKPRSSVDTPPR 399
Db 477 VEIAKQALQNELEKMKELSLKRRGKDLR-----KSEKKAQQTPIED-DNEDLKQ 526
Qy 400 L-----STPQKGPSTHPKEK-APSEIIEIDLPLYLSTTEMYICRMHQPPSPPLPRLS-- 449
Db 527 LQFVKEEALMRKRAKXIDKEORFEHLQ--KYRS-----FTGDLSLPLPKGAGG 576
Qy 450 --SPKKEETVARCLMPSSVAGETSVLA-----VPSMRDHSVLEPLDPP 491
Db 577 PSTREAEKLRLRL--VBERANITGRKIVLEVENRGLKAEILDLDLRSGEDPFGSSNPLM 633
```

```

QY 492 SULENIDSVFSKRNKLEDEKRRKWDIORIREQ-RILORLOLMTKKKGIQES 547
Db 634 REOSESEISE---LROHLQVEDETELLRRVADLEBQNKRIITAEINKYKXSGHDS 688

RESULT 10
QY 07PX26 PRELIMINARY; PRT; 1126 AA.
ID 07PX26
AC 07PX26;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE AGCP12735 (Fragment).
GN Name=agCG49353; ORFNames=ENSNAG0000019625;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RU Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AAB01009887; EMBL: AAB01359.1; -.
DR HSSP: Q00420; IAMC.
DR InterPro: IPR002110; ANK.
DR Pfam: PF00023; ANK; 10.
DR PRINTS: PR01415; ANKYRIN.
DR PROSITE: PS50088; ANK_REPEAT; 6.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
KM ANK repeat.
FT NON_TER 1 1
FT NON_TER 1126 1126
SQ SEQUENCE 1126 AA; 122606 MW; F6C5DE9891C0CB4 CRC64;

Query Match 6.6%; Score 215; DB 2; Length 1126;
Best Local Similarity 22.4%; Pred. NO. 0.33; Indels 168; Gaps 31;
Matches 154; Conservative 87; Mismatches 277;

QY 8 FKAAAPAGGNPRLDYERAAALGSPEDPGAFAHFLPRHRLKEKGPPLASSQ---- 63
Db 530 YKCCSSRSHSRHRNDREBETRDALPDLALKEVFF-KYAEKKEPGRKPEBEGCKI 587
QY 64 -----GSPADSPACCGKRGGLLPAGAAPQOEESWGSVPLPCPPATKQAGITGG 116
Db 588 ILKQVHSGSSSENDSPANCGGR-----AGPLKFSERRRETHLGL-PVQLQYGAGG 637
QY 117 EPAAAGAG-----CSPRPKQAVLP-----QTGSLVAAAEPTPMAGDKGAA 160
Db 638 FDSAGSLDFLEPTSPRTPPRAEFNIRKESDAKSEKSSDSSTTKKTKYKGTGKSKLA 697
QY 161 SPATASDPAPPLPLPGPPPLPATATAGTLAASEGRWMSKSPGSGGSGASSQAA 220
Db 698 KTKAKSSPSPEEQAKRP-----APVTEVVTQAOVHAPESSMSSTDDDAITATY--- 749
QY 221 CLKQILLQLDLLEQ-----QQQLQAK-----EKIEELKSERDTLL 258
Db 750 -----TIEQRTSRDVEGLSETEBSAKGGLPRVTSKPAWETIRREVDEH---- 794
QY 259 ARIERMRMOVLKKNKERRHLFGQYTEREBETELSEKIK-----LECOPELSE- 310
Db 795 -AVEMTQQQQAARCKSKKSAKSSKSHSSSESSDKASKRAPLPQOLEPPELVEE 853
QY 311 -----TSQTLPPKPFSCGRSGKGNRK-SPGSTERKTPVKVLAPEFSKVTKTPKHSPIK 365
Db 854 RTEPTVSSAQPOGLVUNRGSGGGAPEDEPLIAESLPL--VRE-----EPAPKLE 905
QY 366 EEPGGSISFTVCKRELRSQETPEKPRSSVDTPLPLTPQKSPSTHPKPKAFSSIEDLPY 425
Db 906 GEOKSTPEET-----SESTPSK--SSEETEP-TSVVOEVPEKEBQKPAAPPVVEEPPK 955

```

```

QY 426 LSTTEMYLCRWQPPSPPLPRLRESSPKKEETVARCLMPSSVAGETSYLA-VPSSWRHSHVE 484
Db 956 VSEV-----VPVPEAKAQPEPAKPKRLRSMSWMSVVEETALVPA-----E 1000

QY 485 PLRDPNPDSLENUDDS--VFSKRNKLEDEKRRKWDIORIREQIILORLOLMTKKK 542
Db 1001 PV--PPPLDQAK-VDOTRGPFVSLDGTAEDEDEKKK-----KIKKKPRLKEDEQ 1047

QY 543 GISESEPEVTSFPPEPDVSLMITPPLPVARGPRLPKLT-PONFELPWLDER-SRCRL 600
Db 1048 EQSSSKQDQSGFEPPSPAVNS-----KPTV-FERPHYATLPRAAFSVEEBAASSRP 1100

QY 601 EIQK-----KQTPRTRCK 614
Db 1101 EGKRPQKXAVNMTTQGSIRNRIR 1126

RESULT 11
MSL1 DROME
ID MSL1 DROME STANDARD; PRT; 1039 AA.
AC P50535; O9VJ66;
DT 01-OCT-1996 (Rel. 34, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Male-specific lethal-1 protein.
GN Name=msl-1; ORFNames=CG10385;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93300219; PubMed=7781064; DOI=10.1016/0092-8674(95)90007-1;
RA Kelley R.L., Solovyeva I., Lyman L.M., Richman R., Solovyev V.,
RA Kuroda M.I.;
RT "Expression of msl-2 causes assembly of dosage compensation regulators
RT on the x chromosomes and female lethality in Drosophila.";
RL Cell 81:867-877(1995).
RN [2]
RP SEQUENCE OF 85-1039 FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93314941; PubMed=8325488;
RA Palmer M.J., Meynert V.A., Richman R., Manning J.E., Kuroda M.I.,
RA Lucchesi J.C.;
RT "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster
RT encodes a novel protein that associates with the x chromosome in
RT males.";
RL Genetics 134:545-557(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731112; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananthides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton R.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotcler P.,
RA Burris K.C., Busam D.A., Butler H., Cadietu E., Center A., Chandra I.,
RA Cherry J.M., Casale S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegwan C.,

```

RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodish C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro F., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulyov G., Milshina N.V., Mobarry C., Morris J., Mosher A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson S.D., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye Y., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP GENOME REANNOTATION.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Betencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RL systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [5]
 RP INTERACTION WITH TAMO.
 RX PubMed=12653959;
 RA Minakina S., Yang J., Steward R.;
 RT "Tamo selectively modulates nuclear import in *Drosophila*.";
 RL Genes Cells 8:299-310(2003).
 CC -I- FUNCTION: The MSL proteins are essential for elevating
 CC transcription of the single X chromosome in the male (X chromosome
 CC dosage compensation). Msl-1 is a pioneer protein. Msl, msl-1 and
 CC msl-3 are colocalized on the X chromosome. Each of the MSL
 CC proteins requires all the other MSLs for wild-type X-chromosome
 CC binding.
 CC -I- SUBUNIT: Seems to form a tight complex with msl-2. Interacts with
 CC tamo via the nuclear localization signal.
 CC -I- SUBCELLULAR LOCATION: Nuclear; msl-1 is associated with hundreds
 CC of discrete sites along the length of the X chromosome in males
 CC and not in females, and is also associated with 10-20 autosomal
 CC sites in males.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L42514; AAA8918.1; -.
 CC EMBL: AB03660; AAP5368.1; -.
 CC PIR: S52959; S52959.
 CC FLYBASE: FBgn0005617; msl-1.
 CC GO: GO:001456; C:dosage compensation complex (sensu Insecta); NAS.
 CC DR GO: GO:0003682; P:chromatin binding; IDA.
 CC KW Nuclear protein.
 CC DR GO: GO:009047; P:dosage compensation, by hyperactivation of . . .; NAS.
 CC Nucleic acid binding.
 CC FT DOMAIN 1032 1037 Nuclear localization signal.
 CC FT CONFLICT 26 26 R -> G (in Ref. 1).
 CC FT CONFLICT 492 492 S -> L (in Ref. 1).
 CC FT CONFLICT 670 670 M -> I (in Ref. 1).
 CC FT SEQUENCE 1039 AA; 117503 MW; 644C21C0390A2750 CRC64;

Query Match 6 6%; Score 213; DB 1; Length 1039;
 Best Local Similarity 20.9%; Pred. No. 0.37; Indels 118; Gaps 18;
 Matches 99; Conservative 85; Mismatches 172;
 QY 181 PLPAATAGTAAAGGGRKSRKSPPLGGGSSGSSQAACQKQLL-----LQDLIE 234
 DB 569 PHLPRVAAPKTKTSSRSTLPK-----NTADIKAPAKQVIANNGSTKQTDVPK 621
 QY 235 QQQQQQLQAKKE-----IEELKSEBDTLARIEMRERMLQVKDKNEKH 280
 DB 622 TORLQVKIRQYMHDPMTGTSAPSDIRKQKQVDPVSTPETYIKSKMLV--NDKTT 679
 QY 281 KLFQGTETETET-----ELSEKIKLECEPELSETQTL-----KPEFCGS 325
 DB 680 ETSQSPDQIDIVEYARKLAHEHKKELLSQSHSQQVTLKKIRERVAATLIYPP----- 733
 QY 326 GKHKRRKSPGSGTERKTPVKKLAPESKTKYTPK---SPIKEPCGSLSEVCKRELR 382
 DB 734 -----SAVVSST---TTPAPTPPTTPPGSTPGQAVTVSSMDQELSAKSKKAEGI- 783
 QY 383 SQETPEKPRSSVDTPPRISTPQK---GPSTHPKKAFFSSEID-----LPYSTT-- 429
 DB 784 --ATPLTPQSNSSVSSTSTIRKTLNCCSPHTYSKATARGKIQSFRATPEYSTRTWE 841
 QY 430 -----EMYLGRMHQPPSPPLPRESSPKKEETVARCLMPSSVAGETSVLAVPSWRH 841
 DB 842 DQFHCNDSEFL-----EEA-----DELDADPSLEIPKRWV 874
 QY 482 SVEPLRDPNPDLLENIDVSFGRKHALEDEKRRKMDIORIEORILQRLQMLMYKK 541
 DB 875 PVPSPSDKIDTFL---LSDATFERRHQKVQKDEVRKCKDAVYMKQIRLEQRLMRBND 931
 QY 542 KGIQSESP-EVTSFPPEPDVSLMTIFPLPVVAGRPPLKTPQ-NFELPWL 593
 DB 932 EVLVADLPRASTFPLPEDIALQFVNEVTVQAGENVNMEARDDFGVPMVD 985
 RESULT 12
 ID 081ZL8 PRELIMINARY; PRT; 1130 AA.
 AC 081ZL8;
 DT 01-MAR-2003 (TREMELREL. 23, Created)
 DT 01-MAR-2003 (TREMELREL. 26, Last sequence update)
 DT 01-MAR-2004 (TREMELREL. 26, Last annotation update)
 DE MNAR.
 GN Name=MNAR;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22317447; PubMed=12415108; DOI=10.1073/pnas.192569699;
 RA Wong C.W., McNally C., Nickbarg E., Komm B.S., Cheekie B.J.;
 RT "Estrogen receptor-interacting protein that modulates its nongenomic
 RT activity-crosstalk with Src/Erk phosphorylation cascade.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14783-14788(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Wong C.-W., McNally C., Nickbarg E., Komm B.S., Cheekie B.J.;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF547989; AA041255.1; -.
 DR InterPro: IPR008938; ARM.
 SQ SEQUENCE 1130 AA; 119622 MW; 96A698015E3D7CF CRC64;
 Query Match 6 6%; Score 212.5; DB 2; Length 1130;
 Best Local Similarity 22.9%; Pred. No. 0.42; Indels 127; Gaps 19;
 Matches 114; Conservative 54; Mismatches 202;
 QY 55 PGPPPLASQSGG-----SPAPSPACGCGKGRGLLPAGAAPG--QOEESWGGS 99
 DB 700 PGPPPLTANHLGLSVGRLVVPRLRLPGPNHAGSNEDPILAPSGPTTTPPDEFEGGR 759

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.J., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010457; AAH10457.2; -
DR InterPro; IPR008938; ARM.
FT NON TER 1
SQ SEQUENCE 1048 AA; 111876 MW; 48C7A91D241AFABE CRC64;

Query Match 6.5%; Score 209.5; DB 2; Length 1048;
Best Local Similarity 23.1%; Pred. No. 0.52;
Matches 115; Conservative 51; Mismatches 204; Indels 127; Gaps 20;

QY 55 PGPPPLASSQGG-----SPAPSPAGCGGKRGULLPAGAAG--QOESWGGS 99
DB 618 PGPPPTANHLGLSVPLGVSPRLPLPGPNHRAGSNEDPLASGTPPTIPDETFFGG 677
QY 100 VPLPCPPPAATKQ-----AGIGGEPMAAGACSPRPKQAVLP1Q 138
DB 678 VRRPAFVHYDKKEASDVESLSDDSDSVIVVEGLPLPPPPSGATPPP----- 728
QY 139 TGSLLVAAKEPTPWADKGAASPAATASD---PAGPPPLPLPGPPPLAATAGT1A 193
DB 729 ----IAPTGPPPT-----ASPPVPAKEPPEELPAAPGLP-PPPPPPVPVGPVTL 774
QY 194 ASBGRKSKRKSPPLGGGGSGAASQAACIKQILLQLDLIEQQOQQLQAKETIELKSE 253
DB 775 PPG-----LVPEGTPGGGPPALEEDLVININSSD--EEEEEEEEEEEEEEEE 825
QY 254 RDTLLARIERMERRMQLVKDNEKERHKLPGQYETEBRETELSEKIKLECOPELSETSQ 313
DB 826 ED-----FBEEDDEBEYFEEEBEHEEPEEFEEEBEGLEEBEEDDEBEELBEV-- 878
QY 314 TLPPKPFSGSGSGKHKRSPFGSTERKTPVKKLAPEFSKVTKTKTPGHSPIKEEPCGSIS 373
DB 879 ----EDLEFGTAG-----GEVEGAPPPPTLPALPPPSPPKVQCPERPFGILL 925
QY 374 E-----TYCKRELRSQ-ETPEKPRSSVTPRPLSTPOKSPSTHPKEKAF 416
DB 926 EVEEPTGEERGAADTAFTLAPEALPQGEVERGESPAAGPPQELVEEBSAP--TLL 983
QY 417 SSEIEDLPYLTSTMYLCKRWHOPPPSPPLPLRESPPKKEETVARCLMPSSVAGTSVLAVP 476
DB 984 EEETED---GSKKV-----QPPPE-TPAEE--METETEALQEKEDDTAAMLA-- 1028
QY 477 SWRDHSVEPLRDPNPSPD 493
DB 1029 DFIDCPDDEKPPPTPE 1045

RESULT 15
Q6NTB6 PRELIMINARY; PRT; 1130 AA.
AC Q6NTB6;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Proline-, glutamic acid-, leucine-rich protein 1.
GN Name=PLP1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Millhys S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.J., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC069058; AAH69058.1; -
DR InterPro; IPR008938; ARM.
FT NON TER 1
SQ SEQUENCE 1130 AA; 119698 MW; 7B0DEE7A198DA9A6 CRC64;

Query Match 6.5%; Score 209.5; DB 2; Length 1130;
Best Local Similarity 23.1%; Pred. No. 0.56;
Matches 115; Conservative 51; Mismatches 204; Indels 127; Gaps 20;

QY 55 PGPPPLASSQGG-----SPAPSPAGCGGKRGULLPAGAAG--QOESWGGS 99
DB 700 PGPPPTANHLGLSVPLGVSPRLPLPGPNHRAGSNEDPLASGTPPTIPDETFFGG 759
QY 100 VPLPCPPPAATKQ-----AGIGGEPMAAGACSPRPKQAVLP1Q 138
DB 760 VRRPAFVHYDKKEASDVESLSDDSDSVIVVEGLPLPPPPSGATPPP----- 810
QY 139 TGSLLVAAKEPTPWADKGAASPAATASD---PAGPPPLPLPGPPPLAATAGT1A 193
DB 811 ----IAPTGPPPT-----ASPPVPAKEPPEELPAAPGLP-PPPPPPVPVGPVTL 856
QY 194 ASBGRKSKRKSPPLGGGGSGAASQAACIKQILLQLDLIEQQOQQLQAKETIELKSE 253
DB 857 PPG-----LVPEGTPGGGPPALEEDLVININSSD--EEEEEEEEEEEEEEEE 907
QY 254 RDTLLARIERMERRMQLVKDNEKERHKLPGQYETEBRETELSEKIKLECOPELSETSQ 313
DB 857 PPG-----LVPEGTPGGGPPALEEDLVININSSD--EEEEEEEEEEEEEEEE 907
QY 254 RDTLLARIERMERRMQLVKDNEKERHKLPGQYETEBRETELSEKIKLECOPELSETSQ 313
DB 908 ED-----FBEEDDEBEYFEEEBEHEEPEEFEEEBEGLEEBEEDDEBEELBEV-- 960
QY 314 TLPPKPFSGSGSGKHKRSPFGSTERKTPVKKLAPEFSKVTKTKTPGHSPIKEEPCGSIS 373
DB 961 ----EDLEFGTAG-----GEVEGAPPPPTLPALPPPSPPKVQCPERPFGILL 1007
QY 374 E-----TYCKRELRSQ-ETPEKPRSSVTPRPLSTPOKSPSTHPKEKAF 416
DB 1008 EVEEPTGEERGAADTAFTLAPEALPQGEVERGESPAAGPPQELVEEBSAP--TLL 1065
QY 417 SSEIEDLPYLTSTMYLCKRWHOPPPSPPLPLRESPPKKEETVARCLMPSSVAGTSVLAVP 476
DB 1066 EEETED---GSKKV-----QPPPE-TPAEE--METETEALQEKEDDTAAMLA-- 1110
QY 477 SWRDHSVEPLRDPNPSPD 493

Db 1111 DFIDCPPDDEKPPPTB 1127

Search completed: March 23, 2005, 15:31:43
Job time : 184 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2005, 15:34:40 ; Search time 179 Seconds
(without alignments)
1756.517 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 614
Sequence: 1 MTRRSADVFAKAAAPAGGNPE.....RSRCRLIEQKQTPHRTCRK 614

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : UniProt_03:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	312	50.8	413	2	068DK7
2	125	20.4	125	2	069Z03
3	73	11.9	616	2	06PDM1
4	73	11.9	616	2	09CXR9
5	70	11.4	370	2	09D5C9
6	70	11.4	463	2	080XS0
7	70	11.4	598	2	08BPN3
8	10	1.6	329	2	07P174
9	10	1.6	857	2	07K0Y2
10	10	1.6	867	2	06AXB4
11	10	1.6	1006	2	09W3B8
12	10	1.6	1289	2	062717
13	10	1.6	1382	2	061374
14	10	1.6	1390	2	09ULH8
15	10	1.6	1392	2	080RT1
16	9	1.5	102	1	Y334_METOJA
17	9	1.5	159	2	08L559
18	9	1.5	274	2	07R0A8
19	9	1.5	288	2	098F90
20	9	1.5	323	1	OLG2_HUMAN
21	9	1.5	353	1	CERB1_BOVIN
22	9	1.5	353	2	06S170
23	9	1.5	436	2	08NSU3
24	9	1.5	577	2	09STP1
25	9	1.5	582	2	06C3N1
26	9	1.5	594	2	09LGE2
27	9	1.5	631	1	HS73_BOVIN
28	9	1.5	652	2	08NAX1
29	9	1.5	652	2	086UD5
30	9	1.5	893	2	08WZ55
31	9	1.5	986	2	086B90

32	9	1.5	992	2	08MTW7	08MTW7 drosophila
33	9	1.5	1102	2	067TE6	067TE6 symbiodace
34	9	1.5	1274	2	08NFR0	08NFR0 homo sapien
35	9	1.5	1513	2	086B91	086B91 drosophila
36	8	1.3	69	2	09TID9	09TID9 lacticocilli
37	8	1.3	83	2	09MT14	09MT14 zea mays (m
38	8	1.3	98	2	063N43	063N43 burkholderi
39	8	1.3	103	2	08U4B2	08U4B2 pyrococcus
40	8	1.3	106	2	0688T7	0688T7 oryza sativ
41	8	1.3	109	1	RLAI_MALZE	P52855 zea mays (m
42	8	1.3	110	2	06Z4H8	06Z4H8 oryza sativ
43	8	1.3	111	2	06K630	06K630 oryza sativ
44	8	1.3	112	1	RLAI_OSCBR	001339 oscitatus br
45	8	1.3	114	1	RLAI_CHICK	P18660 gallus gall
46	8	1.3	117	2	08TUD7	08TUD7 methanosarc
47	8	1.3	119	2	068CS3	068CS3 homo sapien
48	8	1.3	120	2	092ZD6	092ZD6 rhizobium m
49	8	1.3	127	2	09TB30	09TB30 solanum cha
50	8	1.3	130	2	08GUR8	08GUR8 pisin sativ
51	8	1.3	136	1	GLBN_MYCTU	Q10784 mycobacteri
52	8	1.3	142	2	09MTN5	09MTN5 zea mays (m
53	8	1.3	144	2	08TH57	08TH57 methanosarc
54	8	1.3	144	2	08TH58	08TH58 methanosarc
55	8	1.3	144	2	08TH59	08TH59 methanosarc
56	8	1.3	144	2	08TKT4	08TKT4 methanosarc
57	8	1.3	144	2	08TLT2	08TLT2 methanosarc
58	8	1.3	144	2	08TR68	08TR68 methanosarc
59	8	1.3	144	2	0653C7	0653C7 oryza sativ
60	8	1.3	145	2	075185	075185 oryza sativ
61	8	1.3	146	2	09NMA0	09NMA0 homo sapien
62	8	1.3	156	2	08WNG2	08WNG2 homo sapien
63	8	1.3	156	2	06ETP9	06ETP9 oryza sativ
64	8	1.3	156	2	06DGM8	06DGM8 brachydanio
65	8	1.3	160	2	08SQ36	08SQ36 bos taurus
66	8	1.3	160	2	08SQ37	08SQ37 bos taurus
67	8	1.3	160	2	08SQ38	08SQ38 bos taurus
68	8	1.3	162	2	08LN77	08LN77 oryza sativ
69	8	1.3	162	2	09RBM6	09RBM6 pseudomonas
70	8	1.3	163	2	06CE45	06CE45 yarrowia li
71	8	1.3	165	2	073127	073127 treponema d
72	8	1.3	172	2	066EP9	066EP9 yersinia ps
73	8	1.3	172	2	08ZIU6	08ZIU6 yersinia pe
74	8	1.3	173	2	06UIP1	06UIP1 macaca mula
75	8	1.3	174	2	06UIP2	06UIP2 pan troglod
76	8	1.3	175	2	084XP8	084XP8 brassica ra
77	8	1.3	178	1	FXYS_RAT	FXYS_RAT rat
78	8	1.3	178	2	06P9W0	06P9W0 rattus norv
79	8	1.3	188	2	06EB13	06EB13 trititum ae
80	8	1.3	190	2	08TWC6	08TWC6 methanosarc
81	8	1.3	193	2	08SIC1	08SIC1 toridula go
82	8	1.3	194	2	075LY0	075LY0 oryza sativ
83	8	1.3	196	2	022638	022638 zea mays (m
84	8	1.3	197	2	06MW69	06MW69 oryza sativ
85	8	1.3	197	2	07XP34	07XP34 oryza sativ
86	8	1.3	202	2	06ZVZ8	06ZVZ8 homo sapien
87	8	1.3	212	2	084CJ1	084CJ1 streptomyce
88	8	1.3	216	2	0917R0	0917R0 drosophila
89	8	1.3	223	2	07XA56	07XA56 gnetum gnet
90	8	1.3	227	2	084S12	084S12 oryza sativ
91	8	1.3	251	1	06YRM3	06YRM3 oryza sativ
92	8	1.3	252	1	BTE4_HUMAN	09YXK1 homo sapien
93	8	1.3	252	2	P70866	P70866 borrelia bu
94	8	1.3	256	2	09S047	09S047 borrelia bu
95	8	1.3	259	2	09V6R4	09V6R4 drosophila
96	8	1.3	263	1	KD2_HUMAN	Q14681 homo sapien
97	8	1.3	263	1	08SID5	08SID5 pelliciera
98	8	1.3	263	2	09RKP1	09RKP1 streptomyce
99	8	1.3	269	2	08RUV3	08RUV3 oryza sativ
100	8	1.3	272	2	0985D6	0985D6 rhizobium 1

ALIGNMENTS

```
RESULT 1
Q68DK7 PRELIMINARY; PRT; 413 AA.
AC Q68DK7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Hypothetical protein DKFZp686j17211.
GN Name=DKFZp686j17211;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
SQ
SEQUENCE 413 AA; 47603 MW; 29BA8726890AB0BE CRC64;

Query Match 50.8%; Score 312; DB 2; Length 413;
Best Local Similarity 99.8%; Pred. No. 6,4e-283;
Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 202 MKRSPVGGGGGSSGASQAACIKQIILLQDLIEQOQOQOQAKKEIIBELKSEBDTLART 261
DB 1 MKRSPVGGGGGSSGASQAACIKQIILLQDLIEQOQOQOQAKKEIIBELKSEBDTLART 60
QY 262 EMEREMQVLVKKNKNEKERNHLFGQYETEREETELSEKIKLECOPELSETSOQLPPKPS 321
DB 61 EMEREMQVLVKKNKNEKERNHLFGQYETEREETELSEKIKLECOPELSETSOQLPPKPS 120
QY 322 CGBSGKGHRKSPGSGSTERTKTPYKTLAPFSKVKTKTPKHSPIKEBPGSLSTVCREL 381
DB 121 CGBSGKGHRKSPGSGSTERTKTPYKTLAPFSKVKTKTPKHSPIKEBPGSLSTVCREL 180
QY 382 RSQETEKRRSSVDTPEPLSTPOKGPSTHPKKAFFSEIEDLPYLSTTEMYLCRWHP 441
DB 181 RSQETEKRRSSVDTPEPLSTPOKGPSTHPKKAFFSEIEDLPYLSTTEMYLCRWHP 240
QY 442 SPLPLRESSPKKETVARCIMPSSVAGETSVLAVPGRWDHSEPLRDPNSDLLENLDDS 501
DB 241 SPLPLRESSPKKETVARCIMPSSVAGETSVLAVPGRWDHSEPLRDPNSDLLENLDDS 300
QY 502 VPSKRHAKELEDEKRRRRWDIQIRERQRIQLRLMYKKKGIOESEPEVTSFPEPDDV 561
DB 301 VPSKRHAKELEDEKRRRRWDIQIRERQRIQLRLMYKKKGIOESEPEVTSFPEPDDV 360
QY 562 ESIMTPEPLPYVAFGRPLPLTPONFELPWLDRSRORLEIOKKQPHRTCRK 614
DB 361 ESIMTPEPLPYVAFGRPLPLTPONFELPWLDRSRORLEIOKKQPHRTCRK 413

RESULT 2
Q69Z03 PRELIMINARY; PRT; 125 AA.
AC Q69Z03;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Hypothetical protein DKFZp586b1922.
GN Name=DKFZp586b1922;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
SQ
SEQUENCE FROM N.A.
```

```
RC TISSUE=Uterus;
RA The German cDNA Consortium;
RA Koehler K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049450; CAH10734.1; -.
KM Hypothetical protein.
FT NON_TER
FT 125
SQ SEQUENCE 125 AA; 14065 MW; C6BF5B36FE59E03 CRC64;

Query Match 20.4%; Score 125; DB 2; Length 125;
Best Local Similarity 100.0%; Pred. No. 2.6e-108;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 PGSTERTKTPVKKLAPEFSKVKTKTPKHSPIKEBPGSLSTVCRELRSQETPEKRRS 393
DB 1 PGSTERTKTPVKKLAPEFSKVKTKTPKHSPIKEBPGSLSTVCRELRSQETPEKRRS 60
QY 394 VDTPEPLSTPOKGPSTHPKKAFFSEIEDLPYLSTTEMYLCRWHPPPPLPLRESSPKK 453
DB 61 VDTPEPLSTPOKGPSTHPKKAFFSEIEDLPYLSTTEMYLCRWHPPPPLPLRESSPKK 120
QY 454 EETVA 458
DB 121 EETVA 125

RESULT 3
Q6PDM1 PRELIMINARY; PRT; 616 AA.
AC Q6PDM1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE RIKEN cDNA 4121402D02.
GN Name=4121402D02Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
SQ
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Mouse;
RX MEDLINE=2388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J.W., Green E.D., Dickson M.C.,
RA Krzywinski M.I., Skalska U., Smalios D.E., Schmech A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Mouse;
RA Strausberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC058629; AAH58629.1; -.
SQ SEQUENCE 616 AA; 67319 MW; 2679AF230DBCA7D CRC64;
```

Query Match 11.9%; Score 73; DB 2; Length 616;
Best Local Similarity 100.0%; Pred. No. 5,2e-59;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 421 EDLPYSTTEMYLCRMHQPSPPLPRLSSPKKEETVARCLMPSSVAGETSVLAVPSWRD 480
Db 423 EDLPYSTTEMYLCRMHQPSPPLPRLSSPKKEETVARCLMPSSVAGETSVLAVPSWRD 482

Qy 481 HSEVPLRDPNPSD 493
Db 483 HSEVPLRDPNPSD 495

RESULT 4
Q9CXF9 PRELIMINARY; PRT; 616 AA.
AC Q9CXF9
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
DE library, clone:4121402D02 product: hypothetical protein, full insert
DE sequence.
GN Name=4121402D02Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subfraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20530913; PubMed=11076661; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Suni N., Ishii Y., Nishi K., Kiteuna T., Tashiro H., Itoh M.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kaishiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).

Qy 421 EDLPYSTTEMYLCRMHQPSPPLPRLSSPKKEETVARCLMPSSVAGETSVLAVPSWRD 480
Db 423 EDLPYSTTEMYLCRMHQPSPPLPRLSSPKKEETVARCLMPSSVAGETSVLAVPSWRD 482

Qy 481 HSEVPLRDPNPSD 493
Db 483 HSEVPLRDPNPSD 495

RESULT 5
Q9DSC9 PRELIMINARY; PRT; 370 AA.
AC Q9DSC9
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE library, clone:4930463F05 product: hypothetical protein, full insert
DE sequence.
GN Name=4121402D02Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076661; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Suni N., Ishii Y., Nishi K., Kiteuna T., Tashiro H., Itoh M.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kaishiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).

```
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:11617-16301(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashwagi K.,
RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuyama S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multiplexed capillary sequencer.";
RL Genome Res. 10:11757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoaka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Katsukawa T., Kato H.,
RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito K., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Shibata K., Tagami M., Tagawa A., Tanaka T.,
RA Tejima Y., Toyota T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK014966; BAB29868.1; -
RL MGD; MGI:1921276; 4121402D02Rik.
KW Hypothetical protein.
SQ SEQUENCE 370 AA; 43184 MW; 2347FA88F39B5654 CRC64;
Query Match 11.4%; Score 70; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 2.1e-56;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 495 LENVLDSDVFSKRAKLEDEKRRKRDQIRIRBQRIQLRLQRLMYKKKGIOESEPEVTSTF 554
DB 251 LENVLDSDVFSKRAKLEDEKRRKRDQIRIRBQRIQLRLQRLMYKKKGIOESEPEVTSTF 310
QY 555 FPEPDVVESTL 564
DB 311 FPEPDVVESTL 320
RESULT 6
Q80XS0 PRELIMINARY; PRT; 463 AA.
AC Q80XS0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE 4121402D02Rik protein.
GN Name=4121402D02Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
```

```
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Statchenko M., Soares M.B., Bonaldo M.F., Casavert T.L., Scheetz T.E.,
RA Brownstein M.J., Ussid T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettleman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield J.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC043039; AAH43039.1; -
DR EMBL: BC055715; AAH55715.1; -
DR MGD; MGI:1921276; 4121402D02Rik.
SQ SEQUENCE 463 AA; 49206 MW; C52E141065B00CF2 CRC64;
Query Match 11.4%; Score 70; DB 2; Length 463;
Best Local Similarity 100.0%; Pred. No. 2.6e-56;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 203 RKSPLGGGGGSGASGAPACIKQLILDLIEQQQQQLQAKKEIEBELKSERDTLLARIE 262
DB 205 RKSPLGGGGGSGASGAPACIKQLILDLIEQQQQQLQAKKEIEBELKSERDTLLARIE 264
QY 263 RMERRMQLVK 272
DB 265 RMERRMQLVK 274
RESULT 7
Q8BPN3 PRELIMINARY; PRT; 598 AA.
AC Q8BPN3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
DE library, clone: E130302B09 product: hypothetical protein, full insert
DE sequence.
GN Name=4121402D02Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EyeBall;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EyeBall;
RX MEDLINE=21085560; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
```

```

RL Nature 409:685-690(2001).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Matshiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer."
RL Genome Res. 10:1757-1771(2000).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RA Adachi U., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hatanaka T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiroaka T., Hirokane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numasaki R., Ono M., Ohata N., Okazaki Y.,
RA Saito R., Satoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Taya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK053719; BAC35489.1; -
DR MGD; MGI:1921276; 4121402D02Rik.
KW Hypothetical protein.
SQ
SEQUENCE 598 AA; 65100 MW; 9F10EEBA9AA7ACCS CRC64;
Query Match 11.4%; Score 70; DB 2; Length 598;
Best Local Similarity 100.0%; Pred. No. 3.2e-56;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 203 RKSPILGGGGGSSGASQAACIKQILLQLDLEQQQQQOLAKKKEIEIKSERPTLLARIE 262
DB 205 RKSPILGGGGGSSGASQAACIKQILLQLDLEQQQQQOLAKKKEIEIKSERPTLLARIE 264
QY 263 RMERMQLVX 272
DB 265 RMERMQLVX 274
RESULT 8
Q7P174 PRELIMINARY; PRT; 329 AA.
AC Q7P174;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Probable bacteriophage tail fiber protein.

```

```

GN OrderedLocustNames=CV0339;
OS Chromobacterium violaceum.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Chromobacterium.
OX NCBI_TaxID=536;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12472 / DSM 30191;
RX MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;
RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida F.,
RA Alves-Gomes J.A., Andrade E.M., Aratiga U., de Araujo M.F.F.,
RA Astolfi-Filho S., Azevedo V., Baptista A.J., Batist M.A.M.,
RA Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,
RA Bordignon J., Brigido M.M., Brito C.A., Broccoli M., Burty H.A.,
RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
RA Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,
RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
RA Fantinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furian L.R.,
RA Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangelro T.B.,
RA Grattapaglia D., Grissard E.C., Hanna E.S., Jardim S.N., Laurino J.,
RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
RA Madeira H.M.F., Manlio G.P., Maranhao A.Q., Martins W.S.,
RA di Mauro S.M.Z., de Medeiros S.R.B., Meisner R.V., Moreira M.A.M.,
RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
RA Paimo R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Portrich D.P.,
RA Ramalho-Neto C.B., Reis A.M.M., Rigo L.U., Rondinelli E.,
RA Santos E.B.P., Santos F.R., Schneider M.P.C., Senaes H.N.,
RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Umenyi T.,
RA Vettore A., Waesem R., Zaha A., Simpson A.J.G.;
RT "The complete genome sequence of Chromobacterium violaceum reveals
RT remarkable and exploitable bacterial adaptability."
RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
DR EMBL; AE016911; AA058018.2; -
KW Complete proteome.
SQ
SEQUENCE 329 AA; 31478 MW; 3FB035B44CE4064C CRC64;
Query Match 1.6%; Score 10; DB 2; Length 329;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 208 GGGGGSGASS 217
DB 143 GGGGGSGASS 152
RESULT 9
Q7K0Y2 PRELIMINARY; PRT; 857 AA.
AC Q7K0Y2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE LD23217P.
GN Name=CG12075;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phytroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celnik S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

```

DR EMBL: AY069512; AAL39657.1; -.
DR InterPro: IPR001005; Myb_DNA_binding.
DR PROSITE: PS00037; MYB_1; UNKNOWN 1.
SQ SEQUENCE 857 AA; 97778 MW; A1FD62256BE92B5D CRC64;
Query Match 1.6%; Score 10; DB 2; Length 857;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 208 GGGGGSGGASS 217
DB 809 GGGGGSGGASS 818
RESULT 10
Q6AXB4 PRELIMINARY; PRT; 867 AA.
ID Q6AXB4
AC Q6AXB4;
DT 25-OCT-2004 (TEMBLrel. 28, Created)
DT 25-OCT-2004 (TEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TEMBLrel. 28, Last annotation update)
DE Cadps protein.
GN Name=Cadps;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heile F.,
RA Diatchenko L., Marziska K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalka U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Director MGC Project;
RA Submitted (Aug-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL: BC079679; AAT79679.1; -.
DR GO: GO:0005515; F:protein binding; IPI.
DR InterPro: IPR000008; C2.
DR InterPro: IPR008973; C2_CALB.
DR InterPro: IPR010439; PH.
DR InterPro: IPR011036; PH_related.
DR Pfam: PF00168; C2; 1.
DR Pfam: PF06292; DUF1041; 1.
DR Pfam: PF00169; PH; 1.
DR SMART: SM00233; PH; 1.
DR PROSITE: PSS0003; PH_DOMAIN; 1.
SQ SEQUENCE 867 AA; 97154 MW; AC9879E7209DE4E6 CRC64;
Query Match 1.6%; Score 10; DB 2; Length 867;
Best Local Similarity 100.0%; Pred. No. 6.7;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 208 GGGGGSGGASS 217
DB 60 GGGGGSGGASS 69
RESULT 11
Q9W3B8 PRELIMINARY; PRT; 1006 AA.
ID Q9W3B8
AC Q9W3B8; Q8T070;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE CG12075-PA (CG12075-pb).
GN ORFNames=CG12075;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phyloroidae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazek R.G., Zhang M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Andrews-Pfannkoch C., Baldwin D.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwey R.M., Basu N., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downe M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ijogawa K.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemtson J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberg C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster.";
Science 287:2185-2195(2000).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavelly T., Muzny D.M., Nelson C.R.,
RA Paclet J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
melanogaster euchromatic genome sequence.";

RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=2242607; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskaas R.,
RA Patel S., Friese E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celisner S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=2242606; PubMed=12537572;
RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Bergman B.P.,
RA Battencourt B.R., Celisner S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton K., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; ABO03445; AAF6414.3; -.
DR FlyBase; FBgn0030065; CG12075.
DR InterPro; IPR001005; Myb_DNA_binding.
DR InterPro; IPR011036; PH_related.
DR PROSITE; PSS00037; MYB_1; UNKNOWN 1.
SQ SEQUENCE 1006 AA; 114489 MW; 43E364D7CE345A7F CRC64;

Query Match 1.6%; Score 10; DB 2; Length 1006;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 GGGGGSGASS 217
Db 958 GGGGGSGASS 967

RESULT 12
Q62717 PRELIMINARY; PRT; 1289 AA.
AC 062717;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ca2+-dependent activator protein.
GN Name=CAPS;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92386596; PubMed=1516133; DOI=10.1016/0092-8674(92)90310-9;
RA Walent J.H., Porter B.W., Martin T.F.;
RT "A novel 145 kd brain cytosolic protein reconstitutes Ca(2+)-regulated
RT secretion in permeable neuroendocrine cells.";
RL Cell 70:765-775(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97426024; PubMed=9289490; DOI=10.1074/jbc.272.32.19637;
RA Ann K., Kowalchuk J.A., Loyet K.M., Martin T.F.;

RT "Novel Ca2+-binding protein (CAPS) related to UNC-31 required for
RT Ca2+-activated exocytosis.";
RL J. Biol. Chem. 272:19637-19640(1997).
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL; U16802; AAB8635.1; -.
DR PIR; I84505; I84505.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00239; C2; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
SQ SEQUENCE 1289 AA; 146265 MW; 2ADAEA35FB20BCDD CRC64;

Query Match 1.6%; Score 10; DB 2; Length 1289;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 GGGGGSGASS 217
Db 60 GGGGGSGASS 69

RESULT 13
Q61374 PRELIMINARY; PRT; 1382 AA.
AC 061374;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ca2+ dependent activator protein for secretion.
GN Name=Caps;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STAIN=BLAU/c; TISSUE=Brain;
RA Namikawa K., Su Q.N., Toki H., Kiyama H.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL; D86214; BAA13044.1; -.
DR MGI; MGI:1350922; Caps.
DR GO; GO:0005515; F:Protein binding; IPT.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00239; C2; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
SQ SEQUENCE 1382 AA; 155898 MW; EDD721D60CFE1DE CRC64;

Query Match 1.6%; Score 10; DB 2; Length 1382;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 GGGGGSGASS 217
Db 60 GGGGGSGASS 69

RESULT 14
Q9UTU8 PRELIMINARY; PRT; 1390 AA.
AC 09UTU8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

```

DE KIAA1121 protein (Fragment).
GN Name=KIAA1121.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039618; PubMed=10574461;
RA Hirosewa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;
RT "Characterization of cDNA clones selected by the Genemark analysis
RT from size-fractionated cDNA libraries from human brain.";
RU DNA Res. 6:329-336(1999).
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL; AB032947; BAA86435.2; -.
DR GeneW; HGNC:1426; CADPS.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00239; C2; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
FT NON_TER 1
SQ SEQUENCE 1390 AA; 15665 MW; 5DC5501B89227B97 CRC64;

Query Match 1.6%; Score 10; DB 2; Length 1390;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGASS 217
Db 99 GGGGGSGASS 108

RESULT 15
Q80TJ1 PRELIMINARY; PRT; 1392 AA.
AC Q80TJ1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE KIAA1121 protein (Fragment).
GN Name=Cadps; Synonyms=mkKIAA1121;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Nakajima D., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT 11. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RU DNA Res. 10:35-48(2003).
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL; AK12453; BAC65735.1; -.
DR MGD; MG1:1350922; Cadps.
DR GO; GO:0005515; F:protein binding; IPI.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00239; C2; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.

```

```

FT NON_TER 1
SQ SEQUENCE 1392 AA; 156602 MW; 7A1871C210F030E7 CRC64;

Query Match 1.6%; Score 10; DB 2; Length 1392;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGASS 217
Db 97 GGGGGSGASS 106

RESULT 16
Y334_METUA STANDARD; PRT; 102 AA.
ID Y334_METUA
AC Q57780;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hypothetical protein M0334.
GN OrderedLocustNames=M0334;
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Retch C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utecherback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RU Science 273:1058-1073(1996).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U67487; AAB98322.1; -.
DR PIR; F64341; F64341.
DR TIGR; M0334; -.
KM Complete proteome; Hypothetical protein.
FT DOMAIN 60
SQ SEQUENCE 102 AA; 11555 MW; 1C9A45826A0B2A53 CRC64;

Query Match 1.5%; Score 9; DB 1; Length 102;
Best Local Similarity 100.0%; Pred. No. 8; 6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 KEKEIEELK 251
Db 72 KEKEIEELK 80

RESULT 17
Q8L559 PRELIMINARY; PRT; 159 AA.
ID Q8L559
AC Q8L559;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein F0617C02.108 (Hypothetical protein P0446F04.134)

```


DE (Hypothetical protein B1317D11.103).
GN Name=P0617C02.108; Synonyms=B1317D11.103, P0446F04.134;
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
clone: P0617C02.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
clone: P0446F04.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
clone: B1317D11.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004673; BAC10369.1; -
DR EMBL; AP005187; BAD31114.1; -
DR EMBL; AP006186; BAD31932.1; -
DR Gramene; Q81559; -
KW Hypothetical protein.
SQ SEQUENCE 159 AA; 17599 MW; 1AF4798B82F68917 CRC64;
QY
Query Match 1.5%; Score 9; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 439 PPPSPPLPLR 447
142 PPPSPPLPLR 150

RESULT 18
Q7R0A8 PRELIMINARY; PRT; 274 AA.
AC Q7R0A8;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE GIJ_608_56918_56094.
OS Giardia lamblia ATCC 50803.
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
OX NCBI_TaxID=184922;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WB C6;
RA Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,
RT Olsen G.J., Sogin M.L.;
RT "Draft sequence of the Giardia lamblia genome.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC EMBL; AACB0100036; EAA40768.1; -
DR HSSP; P12282; LTM9
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR009036; Moeb.
DR InterPro; IPR007901; Moeb_Moeb.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR000594; Thif domain.
DR Pfam; PF05237; Moeb_Moeb_1.
DR Pfam; PF00899; Thif_1.
SQ SEQUENCE 274 AA; 30420 MW; ABCE1250D2280EBC CRC64;

Query Match 1.5%; Score 9; DB 2; Length 274;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 188 TAGTLAASE 196
Db 204 TAGTLAASE 212

RESULT 19
Q98F90 PRELIMINARY; PRT; 288 AA.
AC Q98F90;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE M13880 protein.
OS OrderedLocustNames=m13880;
OC Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobium loti (Mesorhizobium loti).
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAF303099; Pubmed=11214968;
RX MEHLIN=21082930; Pubmed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Matsumoto T., Itoh K., Ishikawa A., Kawashima K., Yamada T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsunaga M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimoto S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003003; BAB50677.1; -
KW Complete proteome.
SQ SEQUENCE 288 AA; 29542 MW; 853145F5541A15B7 CRC64;
QY
Query Match 1.5%; Score 9; DB 2; Length 288;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 155 KAAAPAGG 163
9 KAAAPAGG 17

RESULT 20
OIG2 HUMAN STANDARD; PRT; 323 AA.
AC OIG2 HUMAN
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Oligodendrocyte transcription factor 2 (Olig2) (Basic helix-loop-
helix protein class B 1) (Protein kinase C-binding protein RACK17)
DE (Protein kinase C binding protein 2)
GN Name=OIG2; Synonyms=BHLH1, PRKCBP2, RACK17;
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hippocampus;
RA Kuroda S., Tokunaga C., Kiyohara Y., Konishi H., Kikkawa U.;
RT "Protein kinase C-binding protein.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20202664; Pubmed=10737801; DOI=10.1073/pnas.97.7.3497;
RA Wang J., Uani-Sait S.N., Escalon E.A., Carroll A.J., de Jong P.J.,
RA Kirsch I.R., Aplan P.D.;

RT "The t(14;21)(q11.2;q22) chromosomal translocation associated with T-
 RT cell acute lymphoblastic leukemia activates the BHLH1 gene.",
 RT Proc. Natl. Acad. Sci. U.S.A. 97:3497-3502(2000).
 [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heile F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.T., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
 RA Raha S.S., Loquellano N.A., Peters G.D., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.",
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [4]
 RP TISSUE SPECIFICITY.
 RX Marie Y., Sanson M., Mokhtari K., Leuraud P., Kujas M.,
 RA Delacrette J.-Y., Poirier J., Zalc B., Hoang-Xuan K.;
 RT "OLIG2 as a specific marker of oligodendroglial tumour cells.",
 RT Lancet 358:298-300(2001).
 [5]
 RP TISSUE SPECIFICITY.
 RX PubMed=11526205; DOI=10.1073/pnas.181340798;
 RA Lu Q.R., Park J.K., Noll E., Chan J.A., Alberta J.A., Yuk D.-I.,
 RA Alzamora M.G., Louis D.N., Stiles C.D., Rowitch D.H., Black P.M.;
 RT "Oligodendrocyte lineage genes (OLIG) as molecular markers for human
 RT glial brain tumors.",
 RT Proc. Natl. Acad. Sci. U.S.A. 98:10851-10856(2001).
 CC -I- FUNCTION: Required for oligodendrocyte and motor neuron
 CC specification in the spinal cord, as well as for the development
 CC of somatic motor neurons in the hindbrain. Cooperates with OLIG1
 CC to establish the PMN domain of the embryonic neural tube.
 CC Antagonist of V2 interneuron and of NKX2-2-induced V3 interneuron
 CC development (By similarity).
 CC -I- SUBUNIT: Interacts with NKX2-2 (By similarity).
 CC -I- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. The NLS contained
 CC in the bHLH domain could be masked in the native form and
 CC translocation to the nucleus could be mediated by interaction
 CC either with class E bHLH partner protein or with NKX2-2 (By
 CC similarity).
 CC -I- TISSUE SPECIFICITY: Expressed in the brain, in oligodendrocytes.
 CC Strongly expressed in oligodendroglomas, while expression is weak
 CC to moderate in astrocytomas. Expression in glioblastomas highly
 CC variable.
 CC -I- INDUCTION: By SHH. Also induced by NKX6-1 in the developing spinal
 CC cord, but not in the rostral hindbrain (By similarity).
 CC -I- DOMAIN: The bHLH is essential for interaction with NKX2-2 (By
 CC similarity).
 CC -I- DISEASE: Involved in a form of T-cell acute lymphoblastic leukemia
 CC (T-ALL) through a chromosomal translocation t(14;21)(q11.2;q22)
 CC which involves OLIG2 and T-cell receptor alpha chain (TCRA) genes.
 CC -I- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; U48250; AAC72247.1; ALT_INIT.
 CC EMBL; AF221520; AAF61215.1; ALT_INIT.
 CC EMBL; BC047511; AAF47511.1; -.
 CC Genew; HGNC:9398; OLIG2.
 CC MIM; 606386; -.
 CC InterPro; IPR001092; HLH_baaic.
 CC Pfam; PFO0010; HLH; 1.
 CC PROSITE; PS50888; HLH; 1.
 CC Chromosomal translocation; Developmental protein; DNA-binding;
 CC Nuclear protein; Proto-oncogene; Transcription regulation.
 CC FT DNA_BIND 109 120 Basic motif
 CC FT DOMAIN 121 162 Helix-loop-helix motif.
 CC FT DOMAIN 77 88 Poly-Ser.
 CC FT DOMAIN 206 212 Poly-Ala.
 CC FT DOMAIN 225 237 Poly-Ala.
 CC FT DOMAIN 266 270 Poly-Ala.
 CC FT DOMAIN 273 280 Poly-Gly.
 CC FT CONFLICT 150 150 A -> T (in Ref. 3).
 CC SQ SEQUENCE 323 AA; 32384 MW; 0EC9223961062509 CRC64;
 QY 208 GGGGSGSGAS 216
 Db 274 GGGGSGSGAS 282
 Query Match 1.5%; Score 9; DB 1; Length 323;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 21
 CEEA BOVIN STANDARD; PRT; 353 AA.
 ID CEEA BOVIN
 AC 002754;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE CCAAT/enhancer binding protein alpha (C/EBP alpha).
 GN Name=CEBPA;
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovinae; Bos.
 ON NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Japanese black;
 RX MEDLINE=97060678; PubMed=8904724;
 RA Taniguchi Y., Sasaki Y.;
 RT "Nucleotide sequence of bovine C/EBP alpha gene.",
 RL J. Anim. Sci. 74:2554-2554(1996).
 CC -I- FUNCTION: C/EBP is a DNA-binding protein that recognizes two
 CC different motifs: the CCAAT homology common to many promoters and
 CC the enhanced core homology common to many enhancers (By
 CC similarity).
 CC -I- SUBUNIT: Binds DNA as a dimer and can form stable heterodimers
 CC with C/EBP beta and gamma (By similarity).
 CC -I- SUBCELLULAR LOCATION: Nuclear.
 CC -I- SIMILARITY: Belongs to the bZIP family. C/EBP subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; D82984; BAA20095.1; -.
 CC HSSP; P05554; INMO.
 CC InterPro; IPR004827; TF_bZIP.
 CC SMART; SM00338; BRU2; 1.

```

DR PROSITE; PSS0217; BZIP. 1.
KW Activator; DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 91 98 Poly-Ala.
FT DOMAIN 181 188 Poly-Pro.
FT DOMAIN 262 267 Poly-Gly.
FT DNA BIND 281 301 Basic motif.
FT DOMAIN 312 340 Leucine-zipper.
SQ SEQUENCE 353 AA; 37218 MW; 48CDPB726F29C310 CRC64;

Query Match 1.5%; Score 9; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 9; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

Dy 9 KAAAPAG 17
Db 92 KAAAPAG 100

RESULT 22
06S170 PRELIMINARY; PRT; 353 AA.
AC 06S170;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE C/EBP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Kang M.-U., Jeong Y.H., Lee S.M., Park H.Y., Yang J.M., Mun S.C.,
RA Yoon D.H.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the bZIP family.
DR EMBL; AY458599; AAR24284.1; -.
DR HSSP; P28033; 1CI6.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR004827; TF_bZIP.
DR SMART; SM00338; BRZ. 1.
DR PROSITE; PSS0217; BZIP. 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 353 AA; 37155 MW; 87E7A6929CA8625E CRC64;

Query Match 1.5%; Score 9; DB 2; Length 353;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 9; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

Dy 9 KAAAPAG 17
Db 92 KAAAPAG 100

RESULT 23
08NSU3 PRELIMINARY; PRT; 436 AA.
AC 08NSU3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE FLJ10876 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Brain;

```

```

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong J.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguallano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gamarane P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fehey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Schmeich A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; BC031608; AAH31608.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR00949; ELM2.
DR InterPro; IPR009057; Homeodomain like.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF01448; ELM2. 1.
DR Pfam; PF00249; Myb_DNA_binding; 1.
DR SMART; SM00717; SANT; 1.
KW Nuclear protein.
SQ SEQUENCE 436 AA; 47788 MW; B709C850D704484D CRC64;

Query Match 1.5%; Score 9; DB 2; Length 436;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 9; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

Dy 209 GGGGSGASS 217
Db 27 GGGGSGASS 35

RESULT 24
09STP1 PRELIMINARY; PRT; 577 AA.
AC 09STP1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative proline-rich protein.
GN Name=727E11.90; Synonyms=AT4G27850;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,
RA Barrett B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K.,
RA Schueller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]

```

```
RP SEQUENCE FROM N.A.
RA Leonard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
RA Wewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078579; CAB43393.1; -.
DR EMBL; AL161571; CAB81434.1; -.
DR PIR; T09024; T09024.
SQ SEQUENCE 577 AA; 63586 MW; 7A469A89EDB8332D CRC64;

Query Match 1.5%; Score 9; DB 2; Length 577;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 PLPLPGPPP 182
Db 198 PLPLPGPPP 206

RESULT 25
ID 06C3N1 PRELIMINARY; PRT; 582 AA.
AC 06C3N1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Yarrowia lipolytica chromosome E of strain ClIB99 of Yarrowia
DE lipolytica
GN ORFNames=YALI0B33451g;
OS Yarrowia lipolytica ClIB99.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_Taxid=284591;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ClIB99;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Fringuel L., Aigle M., Antchouard V., Babour A., Barbe V.,
RA Barney S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer E., Catolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Giropi A.,
RA Hantaye F., Hennequin C., Jaumaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicoud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenene D., Tekala F., Wesolowski-Jouvel M., Westhof E., Wirth B.,
RA Zentou-Meyer M., Zivanovic I., Bolocin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ClIB99;
RA Genoscope;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382131; CAG80335.1; -.
DR InterPro; IPR002173; PFKB.
DR PROSITE; PS00583; PFKB KINASRS 1; UNKNOWN 1.
SQ SEQUENCE 582 AA; 61683 MW; A0C1235113C4CDA6 CRC64;

Query Match 1.5%; Score 9; DB 2; Length 582;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGSAS 216
Db 228 GGGGSGSAS 236
```

```
RESULT 26
ID 09IGZ7 PRELIMINARY; PRT; 594 AA.
AC 09IGZ7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to Arabidopsis thaliana chromosome 4 BAC T15B16 (WRKY
DE transcription factor 9).
GN Name=WRKY9;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzaceae; Oryza.
OX NCBI_Taxid=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC PubMed=12447438; DOI=10.1038/nature01184;
RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
RA Wu J., Nilmura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
RA Hishihata S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
RA Ikono M., Itoh S., Itoh T., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K.,
RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshitara R., Yukawa K.,
RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
RA Yano M., Jiang J., Gojobori T.;
RT "The genome sequence and structure of rice chromosome 1.";
RL Nature 420:312-316(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC PubMed=15047897; DOI=10.1104/pp.103.034967;
RX Zhang Z.L., Xie Z., Zou X., Casaretto J., Ho T.H., Shen Q.J.;
RA "A rice WRKY gene encodes a transcriptional repressor of the
RT gibberellin signaling pathway in aleurone cells.";
RL Plant Physiol. 134:1500-1513(2004).
RN [3]
RP SEQUENCE FROM N.A.
RA Zhang Z.-L., Xie Z., Zou X., Casaretto J., Ho T.-h.D., Shen Q.J.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002480; BAA9574.1; -.
DR EMBL; BK005012; DAA05074.1; -.
DR Gramene; O9IGZ7; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 1.
DR PROSITE; PSS0811; WRKY; 1.
SQ SEQUENCE 594 AA; 61071 MW; 7A1D4B5640448789 CRC64;

Query Match 1.5%; Score 9; DB 2; Length 594;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 GAASPAATA 166
Db 565 GAASPAATA 573

RESULT 27
ID HS73_BOVIN STANDARD; PRT; 631 AA.
AC P34933;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Heat shock 70 kDa protein 3.
```

GN Name=HSP70-3;
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93122795; PubMed=1478667;
 RA Grosz M.D., Womack J.E., Skow L.C.;
 RT "Systemic conservation of HSP70 genes in cattle and humans.";
 RL Genomics 14:863-868(1992).
 CC - FUNCTION: In cooperation with other chaperones, Hsp70s stabilize
 CC preexistent proteins against aggregation and mediate the folding
 CC of newly translated polypeptides in the cytosol as well as within
 CC organelles. These chaperones participate in all these processes
 CC through their ability to recognize nonnative conformations of
 CC other proteins. They bind extended peptide segments with a net
 CC hydrophobic character exposed by polypeptides during translation
 CC and membrane translocation, or following stress-induced damage.
 CC - SIMILARITY: Belongs to the heat shock protein 70 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL: L10428; AAA30569.1; -
 CC HSPB; P19120; 3HSC.
 CC InterPro: IPR001023; Hsp70.
 CC Pfam: PF00012; HSP70_1
 CC PRINTS: PR00301; HEATSHOCK70.
 CC ProDom: PD000089; Hsp70; 1.
 CC TIGRfam: TIGR01991; HscA; 1.
 CC PROSITE: PS00297; HSP70_1; 1.
 CC PROSITE: PS00329; HSP70_2; 1.
 CC PROSITE: PS01036; HSP70_3; 1.
 CC ATP-binding; Chaperone; Heat shock; Multigene family.
 KW ATP-binding; Chaperone; Heat shock; Multigene family.
 SQ SEQUENCE 631 AA; 69199 MW; 01ACA20600C9322F CRC64;
 QY 208 GGGGGSGAS 216
 Db 614 GGGGGSGAS 622
 Query Match 1.5%; Score 9; DB 1; Length 631;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 28
 08NMA1 PRELIMINARY; PRT; 652 AA.
 ID 08NMA1;
 AC 08NMA1;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein FLJ35713 (Gametogenetin, isoform a).
 GN Name=GN;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Oca T., Suzuki Y., Nishikawa T., Otsubo T., Sugiyama T., Irie R.,
 RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Odayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,

RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shitatori A.,
 RA Sudo H., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y., Ishida S.,
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotta T., Kusano J.,
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imase N.,
 RA Masashiro K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
 RA Motiya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kunagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Oktani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan B., Moore T., Max S.T., Wang J., Hsieh F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Tothyluki S., Cavarni P., Prange C.,
 RA Raba S.S., Lomellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buttefield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Strauberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK93032; BAC04025.1; -
 DR EMBL: BC035008; AF035008.1; -
 DR GO: GO:0016021; C: integral to membrane; ISS.
 DR GO: GO:0005635; C: nuclear membrane; ISS.
 DR GO: GO:0005515; F: protein binding; ISS.
 DR GO: GO:0007276; P: gametogenesis; ISS.
 DR GO: GO:0006461; P: protein complex assembly; ISS.
 DR GO: GO:0008104; P: protein localization; ISS.
 SQ SEQUENCE 652 AA; 66698 MW; 94EB8A5409FA7647 CRC64;
 QY 208 GGGGGSGAS 216
 Db 562 GGGGGSGAS 570
 Query Match 1.5%; Score 9; DB 2; Length 652;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
RESULT 29
Q86U5      PRELIMINARY;      PRT;      652 AA.
AC Q86U5;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Gametogenetin protein 1a.
GN Name=GN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22602663; PubMed=12574169; DOI=10.1074/jbc.M211023200;
RA Lu B., Bishop C.E.;
RT "Mouse GGN1 and GGN3, two germ cell-specific proteins from the single
RT gene Ggn, interact with mouse POG and play a role in
RT spermatogenesis."
RL J. Biol. Chem. 278:16289-16296(2003).
DR EMBL; AF538035; AAP31500.1; -.
DR GO; GO:0016021; C:integral to membrane; ISS.
DR GO; GO:0005635; C:nuclear membrane; ISS.
DR GO; GO:0005515; F:protein binding; ISS.
DR GO; GO:0007276; P:gametogenesis; ISS.
DR GO; GO:0006461; P:protein complex assembly; ISS.
DR GO; GO:0008104; P:protein localization; ISS.
SQ SEQUENCE 652 AA; 66869 MW; 1FB8FF7F54FD46A9 CRC64;

Query Match      1.5%; Score 9; DB 2; Length 652;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      208 GGGGGSGAS 216
Db      562 GGGGGSGAS 570

RESULT 30
Q8WZS5      PRELIMINARY;      PRT;      893 AA.
AC Q8WZS5;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Related to tpa inducible protein.
GN Name=B8J21.120;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCB1_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hohnsels J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL669989; CAD21098.1; -.
SQ SEQUENCE 893 AA; 95708 MW; F0E754CEA9398093 CRC64;

Query Match      1.5%; Score 9; DB 2; Length 893;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      208 GGGGGSGAS 216
Db      794 GGGGGSGAS 802
```

```
RESULT 31
Q86B90      PRELIMINARY;      PRT;      986 AA.
AC Q86B90;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE CG31247-PC.
GN Name=tinc; ORFNames=CG31247;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCB1_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Adamatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blaziel R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Aghayani A., An H.U., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolehakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cwley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mayhew A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostel C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodgett M., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Ceiniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Murry D.M., Nelson C.R.,
RA Paclebo J.M., Park S., Pfeiffer B.D., Richards S., Sodegren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
```

RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 a genomics perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Battencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF003719; AAC01574.1; -;
 SQ FlyBase; FBgn0038554; tinc.
 SO SEQUENCE 986 AA; 107116 MW; 91DBA76579821327 CRC64;
 Query Match 1.5%; Score 9; DB 2; Length 986;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 235 QQQQQQLOAK 243
 Db 40 QQQQQQLOAK 48
 RESULT 32
 Q8MTW7 PRELIMINARY; PRT; 992 AA.
 AC Q8MTW7; Q9VEB7;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 25-OCT-2004 (TREMblrel. 28, Last annotation update)
 DE Transmembrane protein TINCAR (CG31247-PB).
 GN Name=tinc; ORFNames=CG31247;
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22505163; PubMed=12617821; DOI=10.1016/S0925-4773(02)00387-8;
 RA Hirota Y., Sawamoto K., Okano H.,
 RT "tinc encodes a novel transmembrane protein expressed in the Timan-
 expressing cardioblasts of *Drosophila*.";
 RL Gene Expr. Patterns 2:323-327(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gale R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H., Blazer V., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter B.G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berens P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Burtis K.C., Bueam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.B., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Gloder A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honick J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibbegan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lesko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yen R.F., Zaveri U.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
 melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 a genomics perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Battencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF003719; AAC01574.1; -;
 DR EMBL; AF003719; AAC01574.1; -;
 DR Intact; Q8MTW7; -;

```

DR FLYBase; FBgn0038554; tinc.
DR GO: GO:0016021; C: integral to membrane; IEA.
KW Transmembrane.
SQ SEQUENCE 992 AA; 107875 MW; 79EB30647E2A5C3C CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 2; Length 992;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 QQQQQQLQAK 243
DB 40 QQQQQQLQAK 48

RESULT 33
Q67TE6 PRELIMINARY; PRT; 1102 AA.
ID Q67TE6
AC Q67TE6;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Putative serine protease.
GN ORFNames=STR62;
OS Symbiodacterium thermophilum.
OC Bacteria; Actinobacteria; Symbiodacterium.
OX NCBI_TaxId=2734;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAM14863;
RA Ueda K., Yamashita A., Ishikawa J., Shimada M., Wateuji T.,
RA Morimura K., Ikeda H., Hattori M., Bepu T.;
RT "Complete genome sequence of an uncultured bacterium Symbiodacterium
thermophilum."
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP006840; BAD39047.1; -
DR GO: GO:0008233; F: Peptidase activity; IEA.
DR InterPro; IPR000209; Pept_S8_S53.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PRO0723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 1102 AA; 117264 MW; B5A04B0677A86CE2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 2; Length 1102;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 GGSPPAPSPA 72
DB 472 GGSPPAPSPA 480

RESULT 34
Q8NFR0 PRELIMINARY; PRT; 1274 AA.
ID Q8NFR0
AC Q8NFR0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Calcium-dependent activator protein for secretion protein.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;

RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22546704; PubMed=12659812; DOI=10.1016/S0888-7543(02)00040-X;
RA Cisternas F.A., Vincent J.B., Scherer S.W., Ray P.N.;
RT "Cloning and characterization of human CAPS and CAPS2, new members
of the Cap2+ dependent activator for secretion protein family."
RL Genomics 81:279-291(2003).

```

```

CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL; AF458662; AAM61861.1; -.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00239; C2; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
SQ SEQUENCE 1274 AA; 144075 MW; 5E68818D8B956AC0 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 2; Length 1274;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 GGGGSGGASS 217
DB 63 GGGGSGGASS 71

RESULT 35
Q86B91 PRELIMINARY; PRT; 1513 AA.
ID Q86B91
AC Q86B91;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG31247-PA (CG31247-pd).
GN Name=tinc; ORFNames=CG31247;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;

RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Adair J.F., Agdayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Bakendale J., Bayraktaroglu I., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ileguam J.A.,
RA Jalali M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazolo M., Plitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheibel F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Styckas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinlock G.M., Weissenbach J.,
RA Williams S.M., Woodgett, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```


DT 25-OCT-2004 (TReMBLrel. 28, Created)
 DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
 DE Putative exported protein.
 GN ORFNames=BPS0454;
 OS Burkholderia pseudomallei K96243.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Burkholderia.
 OK NCBI_TaxID=272560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K96243;
 RX PubMed=15377794;
 RA Holden M.T.G., Tibball R.W., Peacock S.J., Cerdano-Tarraga A.M.,
 RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.,
 RA Bentley S.D., Sebahia M., Thomson N.R., Bason N., Beacham I.R.,
 RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
 RA Chillingworth T., Cronin A., Crosset B., Davis P., Desnazer D.,
 RA Fellwell T., Fraser A., Hance Z., Hauser H., Holroyd S., Jagsels K.,
 RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
 RA Rabbittovitch E., Rutherford K., Sanders M., Simmonds M.,
 RA Songsvilva S., Stevens K., Tumapa S., Vesaratchaveat M.,
 RA Whitehead S., Yeats C., Barrett B.G., Oyston P.C.F., Parkhill J.,
 RT "Genomic plasticity of the causative agent of melioidosis,
 RT Burkholderia pseudomallei".
 RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
 DR EMBL, BX571966; CAH37910.1; -.
 SQ SEQUENCE 98 AA; 9805 MW; FCE73BBFF17BED24 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 96;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 KAAAPAG 16
 Db 71 KAAAPAG 78
 RESULT 39
 ID 08U4B2 PRELIMINARY; PRT; 103 AA.
 AC 08U4B2;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE ATPase subunit B.
 GN OrderedLocNames=PF0176;
 OS Pyrococcus furiosus.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OK NCBI_TaxID=2261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the Pyrococcus furiosus genome."
 RT Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AE010144; AAL80300.1; -.
 KW Complete proteome.
 SQ SEQUENCE 103 AA; 11961 MW; FEEA2B24137E3351 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 103;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 EKIEELK 251
 Db 75 EKIEELK 82
 RESULT 40
 ID 0688T7 PRELIMINARY; PRT; 106 AA.
 Q688T7

AC 0688T7;
 DT 25-OCT-2004 (TReMBLrel. 28, Created)
 DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
 DE Hypothetical protein OSUJBa0017N18.14.
 GN Name=OSUJBa0017N18.14;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriaristidae; Oryzaceae; Oryza.
 OK NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chow T.-Y., Hsing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M.,
 RA Chao Y.-T., Chang S.-J., Chen H.-C., Chen S.-K., Chen T.-R.,
 RA Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y., Hsiao S.-H.,
 RA Hsiung J.-N., Hsu C.-H., Huang J.-J., Kau P.-I., Lee M.-C.,
 RA Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W.,
 RA Wu H.-P., Shaw J.-F.,
 RT "Oryza sativa BAC OSUJBa0017N18 genomic sequence."
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AC120988; AAU10701.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 106 AA; 10475 MW; 94E6B2E5836A2231 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 106;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 GGGGSGA 215
 Db 28 GGGGSGA 35

RESULT 41
 ID RLAI_MAIZE STANDARD; PRT; 109 AA.
 AC P52855; O24414;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE 60S acidic ribosomal protein P1 (L12).
 GN Name=RPPIA;
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoidae; Andropogoneae; Zea.
 OK NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. W22;
 RA Hamilton D.A., Turcich M.P., Bokhari-Riza A., Mascarenhas J.P.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-21.
 RC STRAIN=cv. B73; TISSUE=Ear;
 RX MEDLINE=97422884; PubMed=9276949; DOI=10.1104/pp.114.4.1293;
 RA Bailey-Serres J., Vangala S., Szick K., Lee C.H.;
 RT "Acidic phosphoprotein complex of the 60S ribosomal subunit of maize
 RT seedling roots. Components and changes in response to flooding."
 RL Plant Physiol. 114:1293-1305(1997).
 CC - FUNCTION: Plays an important role in the elongation step of
 CC protein synthesis (By similarity).
 CC - SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit
 CC (By similarity).
 CC - PTM: Phosphorylated.
 CC - PTM: The N-terminus is blocked.
 CC - SIMILARITY: Belongs to the ribosomal protein L12P family.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL, outstation on
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL: U00147; AAA91168.1; -.
DR EMBL: U62752; AAB71079.1; -.
DR PIR; T02039; T02039.
DR PIR; T02716; T02716.
DR MaizeDB; 84941; -.
DR InterPro; IPR001813; Ribosomal_60S.
DR InterPro; IPR001853; Ribosomal_P2.
DR Pfam; PF00428; Ribosomal_60S; 1.
DR PRINTS; PR00456; RIBOSOMALP2.
KW Direct protein sequencing; Phosphorylation; Ribosomal protein.
FT CONFLICT 9 9 R -> T (in Ref. 2).
SQ SEQUENCE 109 AA; 11096 MW; 1BFEC834391F080 CRC64;

Query Match
Best Local Similarity 1.3%; Score 8; DB 1; Length 109;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17
Db 72 AAAAPAG 79

RESULT 42
ID Q6Z4H8 PRELIMINARY; PRT; 110 AA.
AC Q6Z4H8;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Hypothetical protein OSJNB0003E08.35 (Hypothetical protein
DE OSJNB0049108.3).
GN Name=OSJNB0003E08.35; Synonyms=OSJNB0049108.3;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OC NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
RT clone:OSJNB0003E08.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
RT clone:OSJNB0049108.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005173; BAC83862.1; -.
DR EMBL; AP005124; BAD30875.1; -.
KW Hypothetical protein.
SQ SEQUENCE 110 AA; 11134 MW; 50D4C9B81C0DB09C CRC64;

Query Match
Best Local Similarity 1.3%; Score 8; DB 2; Length 110;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
Db 99 GGGGGSGA 106

RESULT 43
ID Q6K630 PRELIMINARY; PRT; 111 AA.
AC Q6K630;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

```

```

DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein OJ1004_A05.25.
GN Name=OJ1004_A05.25;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OC NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005286; BAD19691.1; -.
KW Hypothetical protein.
SQ SEQUENCE 111 AA; 10767 MW; 1C1784CC5998E6D6 CRC64;

Query Match
Best Local Similarity 1.3%; Score 8; DB 2; Length 111;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
Db 53 GGGGGSGA 60

RESULT 44
ID RLAL_OSCBR STANDARD; PRT; 112 AA.
AC 001359;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE 60S acidic ribosomal protein P1 (Ribosomal protein RPL-21).
GN Name=rpl-21;
OS Osecheus brevesophaga.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Rhabditinae; Osecheus.
OC NCBI_TaxID=57871;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CEM1;
RX MEDLINE=97420754; PubMed=9275196; DOI=10.1073/pnas.94.18.9751;
RA Evans D., Zorilo D.A.R., Macmorris M., Winter C.E., Lea K.,
RA Blumenthal T.;
RT "Operons and S12 trans-splicing exist in nematodes outside the genus
RT Caenorhabditis.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9751-9756(1997).
CC -1- FUNCTION: Plays an important role in the elongation step of
CC protein synthesis.
CC -1- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit
CC (By similarity).
CC -1- SIMILARITY: Belongs to the ribosomal protein L12P family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL; U90830; AAB71726.1; -.
DR PIR; T10267; T10267.
DR InterPro; IPR001813; Ribosomal_60S.
DR Pfam; PF00428; Ribosomal_60S; 1.
KW Ribosomal protein.
SQ SEQUENCE 112 AA; 11246 MW; 30FD1B8167C958CC CRC64;

Query Match
Best Local Similarity 1.3%; Score 8; DB 1; Length 112;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17

```

```

Db          |||||
            75 AAAAPAG 82

RESULT 45
RL1 CHICK
ID RL1 CHICK STANDARD; PRT; 114 AA.
AC P1660;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE 60S acidic ribosomal protein P1.
GN Name=RP16P1;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89064804; PubMed=3197716;
RA Ferris J.A., Reinach F.C.;
RT "The complete sequence of a chicken-muscle cDNA encoding the acidic
RT ribosomal protein P1."
RL Eur. J. Biochem. 177:513-516(1988).
CC -1- FUNCTION: Plays an important role in the elongation step of
CC protein synthesis.
CC -1- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit.
CC -1- SIMILARITY: Belongs to the ribosomal protein L12P family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X13876; CAJ3280.1; -.
DR PIR; S02029; RSCH2E.
DR InterPro; IPR001813; Ribosomal_60S.
DR InterPro; IPR001859; Ribosomal_P2.
DR Pfam; PF00428; Ribosomal_60S; 1.
DR PRINTS; PR00456; RIBOSOMA1P2.
DR Phosphorylation; Ribosomal protein.
FT MOD_RES 101 101 Phosphoserine (By similarity).
FT MOD_RES 104 104 Phosphoserine (By similarity).
SQ SEQUENCE 114 AA; 11477 MW; C53CAF5F9389C363 CRC64;

Query Match 1.3%; Score 8; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17
Db 71 AAAAPAG 78

RESULT 46
Q8TUD7
ID Q8TUD7 PRELIMINARY; PRT; 117 AA.
AC Q8TUD7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein MA3848.
GN OrderedLocustNames=MA3848;
OS Methanococcus acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC SMRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Annot D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario B.C.,
RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011096; AAM07199.1; -.
DR InterPro; IPR009057; Homeodomain_1ike.
KM Complete proteome; Hypothetical protein.
SQ SEQUENCE 117 AA; 13331 MW; 651F78012EC8D4C3 CRC64;

```

```

Query Match 1.3%; Score 8; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 EIBELKSE 253
Db 83 EIBELKSE 90

```

```

RESULT 47
Q68CS3
ID Q68CS3 PRELIMINARY; PRT; 119 AA.
AC Q68CS3;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein DKFZp781i1119 (Hypothetical protein
DE DKFZp781G1119).
GN Name=DKFZp781i1119; Synonyms=DKFZp781G1119;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RG The German cDNA Consortium;
RA Koehler K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR749805; CAH1865.1; -.
DR EMBL; CR749246; CAH18102.1; -.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KM Hypothetical protein.
SQ SEQUENCE 119 AA; 13023 MW; C51B62C7EB93B020 CRC64;

```

```

Query Match 1.3%; Score 8; DB 2; Length 119;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 416 FSSIEDL 423
Db 51 FSSIEDL 58

```

RESULT 48
Q922D6 PRELIMINARY; PRT; 120 AA.
ID Q922D6
AC Q922D6
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN ORFNames=SM1018;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymA.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021.
RX MEDLINE=2136509; PubMed=11481432; DOI=10.1073/pnas.161294798;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Hutzar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yen K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9886(2001).
DR EMBL: AE007245; AAK65210.1; -.
DR PIR: H95330; H95330.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 120 AA; 12482 MW; A43950EB49FE5C8C CRC64;

Query Match 1.3%; Score 8; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 LLLQLDLI 233
Db 104 LLLQLDLI 111

RESULT 49
Q9TB30 PRELIMINARY; PRT; 127 AA.
ID Q9TB30
AC Q9TB30
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Fl-ATPase alpha subunit (Fragment).
OS Solanum chacoense (Chaco potato).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4108;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L25;
RX Yang Q., Qin X., Liu D.T., Laublin G., Morse D., Cappadocia M.;
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF181565; AAD56026.1; -.
DR GO: GO:0005739; C:mitochondrion; IEA.
DR GO: GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0046933; F:hydrogen-transporting ATP synthase activity. . .; IEA.
DR GO: GO:0046961; F:hydrogen-transporting ATPase activity. . .; IEA.
DR GO: GO:0015986; P:ATP synthesis coupled proton transport; IEA.
KM Mitochondrion.
FT NON_TER 1 1
SQ SEQUENCE 127 AA; 13823 MW; 785E9971CC2DCB78 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 127;

Best Local Similarity 100.0%; Pred. No. 91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
Db 71 AATASDPA 78

RESULT 50
O8GUR8 PRELIMINARY; PRT; 130 AA.
ID O8GUR8
AC O8GUR8
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Thioedoxin h.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3886;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22742423; PubMed=12857849; DOI=10.1104/ps.102.019562;
RA Montichard F., Renard M., Alkhaliloui F., Duval F.D., Machere J.;
RT "Identification and differential expression of two thioedoxin h
RT isoforms in germinating seeds from pea."
RL Plant Physiol. 132:1707-1715(2003).
DR EMBL: AY170651; AAO12855.1; -.
DR HSP: P10599; IWT.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR006662; Thioed.
DR InterPro: IPR005745; Thioedoxin.
DR InterPro: IPR006663; Thioedox_dom2.
DR Pfam: PF00085; Thioedoxin; 1.
DR PRINTS: PR00421; THIOEDOXIN.
DR TIGRFAMs: TIGR01068; thioedoxin; 1.
DR PROSITE: PS00194; THIOEDOXIN; 1.
KW Redox-active center.
SQ SEQUENCE 130 AA; 14550 MW; 4086F65F1BECF5AE CRC64;

Query Match 1.3%; Score 8; DB 2; Length 130;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 IGGEPAHA 121
Db 9 IGGEPAHA 16

RESULT 51
GLEN MYCTU STANDARD; PRT; 136 AA.
ID GLEN MYCTU
AC Q10784;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Hemoglobin-like protein hBN (Flavohemoglobin).
GN Name=g1bN; OrderedLocustNames=RV1542c, MT1594, Mb1569c;
GN ORFNames=MTCY48.23;
OS Mycobacterium tuberculosis,
OS Mycobacterium bovis, and
OS Mycobacterium smegmatis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765, 1772;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,
RA Harris D.E., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III,

RA Tekala F., Badcock K., Basham D., Brown D., Chillingworth T.,
RA Connor R., Davies R.M., Devlin K., Feltevel T., Gentles S., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Krogan J., McLean J., Moule S.,
RA Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RT Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
DOI=10.1128/JB.184.19.5479-5490.2002;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.D., Deboy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,
RA Hickey B.K., Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D.,
RA Salzberg S.L., Delcher A., Uterback T.R., Weidman J.F., Khouri H.M.,
RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,
RA Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RT J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC SPECIES=M.bovis; STRAIN=BCG;
RX MEDLINE=99432214; PubMed=10500158; DOI=10.1073/pnas.96.20.11223;
RA Couture M., Yeh S.R., Wittenberg B.A., Wittenberg J.B., Ouellet Y.,
RA Rouseau D.L., Guertin M.;
RT "A comparative oxygen-binding hemoglobin from Mycobacterium
RT tuberculosis.";
RT Proc. Natl. Acad. Sci. U.S.A. 96:11223-11228(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RA Garner T., Eismeler K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Menseme C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parhill J., Barrett B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=M.smegmatis; STRAIN=LR222;
RA Kumar A., Rawat V.P.S., Das R.H.;
RT "Cloning, expression and characterization of globin like genes of
RT Mycobacterium in insect cells (sf9).";
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=20102674; PubMed=1063862; DOI=10.1074/jbc.275.3.1679;
RA Yeh S.R., Couture M., Ouellet Y., Guertin M., Rouseau D.L.;
RT "A cooperative oxygen binding hemoglobin from Mycobacterium
RT tuberculosis. Stabilization of heme ligands by a distal tyrosine
RT residue.";
RT J. Biol. Chem. 275:1679-1684(2000).
CC -I- FUNCTION: Binds oxygen cooperatively with very high affinity
CC (P50) = 0.013 mmHg at 20 degrees Celsius because of a fast
CC combination (25 microM(-1).s(-1)) and a slow dissociation (0.2 s(-
CC 1)) rate.
CC -I- SUBUNIT: Homodimer.
CC -I- SIMILARITY: Belongs to the protozoan/cyanobacterial globin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, EX842577; CAA98320.1; .

DR EMBL, AF000516; AAK45860.1; .
DR EMBL, AF130980; AAD28758.1; .
DR EMBL, BX248339; CAD96236.1; .
DR EMBL, AJ249386; CAB56291.1; .
DR PIR, C70761; C70761.
DR PDB, 1IDR: X-ray; A/B=1-136.
DR TIGR, MT1594; .
DR TubercuList, RV1542c; .
DR InterPro, IPR001486; Bac_globin.
DR InterPro, IPR009050; Globin_like.
DR Pfam, PF01152; Bac_globin; 1.
DR ProDom, PD004840; Bac_globin; 1.
DR PROSITE, PS01213; GLOBIN_FAM_2; 1.
KW 3D-structure; Complete proteome; Heme; Oxygen transport; Transport.
FT METAL 81 Iron (heme proximal ligand) (Potential).
FT HELIX 3 8
FT TURN 9 10
FT HELIX 15 18
FT TURN 19 20
FT HELIX 21 37
FT TURN 38 38
FT TURN 40 42
FT HELIX 43 46
FT TURN 47 48
FT HELIX 51 65
FT TURN 66 67
FT HELIX 77 81
FT TURN 82 83
FT HELIX 88 104
FT TURN 105 106
FT HELIX 109 119
FT HELIX 120 124
FT TURN 125 125
SQ SEQUENCE 136 AA; 1448 MW; B75D01A5BC064BB CRC64;
QY 28 AALAGPE 35
Db 63 AALAGPE 70
RESULT 52
Q9M7N5 PRELIMINARY; PRT; 142 AA.
ID Q9M7N5
AC Q9M7N5
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE MPPI attachment factor 1.
GN Name=mpfi;
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_Taxid=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99418887; PubMed=10488241;
RA Gundullis F., Pfeiffer N.U., Meier I.;
RT "WAF1, a novel plant protein interacting with matrix attachment region
RT binding protein MPPI, is located at the nuclear envelope.";
RL Plant Cell 11:1755-1768(1999).
DR EMBL, ARI18114; AAF63658.1; .
SQ SEQUENCE 142 AA; 1446 MW; 953A50427E7F4B06 CRC64;
QY Query Match 1.3%; Score 8; DB 1; Length 136;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
10 AAAAPAGG 17

Db 13 AAAAPAG 20

RESULT 53

Q8TH57 PRELIMINARY; PRT; 144 AA.

AC Q8TH57; 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein MA3005 (Hypothetical protein MA3222).
 GN OrderedLocustNames=MA3005, MA3222;
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 NCBI_TaxID=2214;

SEQUENCE FROM N.A.

RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
 RA Galagan J.E., Nisbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Titrrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity."
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AE011002; AA06378.1; -;
 DR EMBL; AE011026; AA06593.1; -;
 DR InterPro: IPR009057; Homeodomain like.
 KW Complete proteome; Hypothetical protein.
 SO SEQUENCE 144 AA; 16139 MW; 13F89C548FDP3665 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 144;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 EISELKSE 253

Db 83 EISELKSE 90

RESULT 54

Q8TH58 PRELIMINARY; PRT; 144 AA.

AC Q8TH58; 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein MA2785 (Hypothetical protein MA3511)
 DE (Hypothetical protein MA4225) (Hypothetical protein MA3375).
 GN OrderedLocustNames=MA2785, MA3375, MA3511, MA4225;
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 NCBI_TaxID=2214;

SEQUENCE FROM N.A.

RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
 RA Galagan J.E., Nisbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Titrrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,

RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity."
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AE010978; AA06163.1; -;
 DR EMBL; AE011060; AA06874.1; -;
 DR EMBL; AE011134; AA07570.1; -;
 DR EMBL; AE011043; AA06744.1; -;
 DR InterPro: IPR009057; Homeodomain like.
 KW Complete proteome; Hypothetical protein.
 SO SEQUENCE 144 AA; 16139 MW; 13F89C525269EB35 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 144;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 EISELKSE 253

Db 83 EISELKSE 90

RESULT 55

Q8TH59 PRELIMINARY; PRT; 144 AA.

AC Q8TH59; 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein MA1650 (Hypothetical protein MA4108).
 GN OrderedLocustNames=MA1650, MA4108;
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 NCBI_TaxID=2214;

SEQUENCE FROM N.A.

RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
 RA Galagan J.E., Nisbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Titrrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity."
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AE010838; AA07508.1; -;
 DR EMBL; AE011123; AA07456.1; -;
 DR InterPro: IPR009057; Homeodomain like.
 KW Complete proteome; Hypothetical protein.
 SO SEQUENCE 144 AA; 16112 MW; 13F89C5247B3DB35 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 144;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 EISELKSE 253

Db 83 EISELKSE 90

RESULT 56

```

08TKT4
ID 08TKT4 PRELIMINARY; PRT; 144 AA.
AC 08TKT4;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, last annotation update)
DE Hypothetical protein MA3311.
GN OrderedlocusNames=MA3311.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RA "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE01097; AAM06320.1; -.
DR InterPro; IPR009057; Homeodomain_like.
DR EMBL; AE011035; AAM06681.1; -.
DR Genome Res. 12:532-542(2002).
DR InterPro; IPR009057; Homeodomain_like.
KW Complete proteome.
SQ SEQUENCE 144 AA; 16093 MW; 061893B25DB3DB35 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 EIEELKSE 253
Db 83 EIEELKSE 90

RESULT 57
08TKT2 PRELIMINARY; PRT; 144 AA.
ID 08TKT2
AC 08TKT2;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, last annotation update)
DE Hypothetical protein MA2946.
GN OrderedlocusNames=MA2946.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RA "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010801; AAM04732.1; -.
KW Complete proteome.
SQ SEQUENCE 144 AA; 16142 MW; A6FC9C5247B3C728 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 EIEELKSE 253
Db 83 EIEELKSE 90

RESULT 59
0653C7 PRELIMINARY; PRT; 144 AA.
ID 0653C7
AC 0653C7;
DT 25-OCT-2004 (TEMBLrel. 28, Created)
DT 25-OCT-2004 (TEMBLrel. 28, last sequence update)
DT 25-OCT-2004 (TEMBLrel. 28, last annotation update)
DE Hypothetical protein P0635G10.19.
GN Name=P0635G10.19;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.

```

```

RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE01097; AAM06320.1; -.
DR InterPro; IPR009057; Homeodomain_like.
KW Complete proteome.
SQ SEQUENCE 144 AA; 16154 MW; 1C0BB678CA0519FC CRC64;

Query Match 1.3%; Score 8; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 EIEELKSE 253
Db 83 EIEELKSE 90

RESULT 58
08TKR68 PRELIMINARY; PRT; 144 AA.
ID 08TKR68
AC 08TKR68;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, last annotation update)
DE Hypothetical protein MA1315.
GN OrderedlocusNames=MA1315.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RA "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010801; AAM04732.1; -.
KW Complete proteome.
SQ SEQUENCE 144 AA; 16142 MW; A6FC9C5247B3C728 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 EIEELKSE 253
Db 83 EIEELKSE 90

RESULT 59
0653C7 PRELIMINARY; PRT; 144 AA.
ID 0653C7
AC 0653C7;
DT 25-OCT-2004 (TEMBLrel. 28, Created)
DT 25-OCT-2004 (TEMBLrel. 28, last sequence update)
DT 25-OCT-2004 (TEMBLrel. 28, last annotation update)
DE Hypothetical protein P0635G10.19.
GN Name=P0635G10.19;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.

```



```

OX NCBI_TaxID=39947;
RN
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, PAC
RT clone: P0635610."
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005396; BAD46090.1; -
KW Hypothetical protein.
SQ SEQUENCE 144 AA; 15070 MW; 1133587D28B7F225 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 GPPPLPLP 178
DB 86 GPPPLPLP 93

RESULT 60
ID Q75L85 PRELIMINARY; PRT; 145 AA.
AC Q75L85;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Putative lOB domain protein.
GN Name=OJ1729_E02.9;
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN
RP SEQUENCE FROM N.A.
RA Chow T.-Y., Hsing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M.,
RA Chao Y.-T., Chang S.-J., Chen S.-K., Chen T.-R.,
RA Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y., Hsiao S.-H.,
RA Heiung J.-N., Hsu C.-H., Huang J.-J., Kau P.-I., Lee M.-C., Leu H.-L.,
RA Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W.,
RA Wu H.-P., Shaw J.-F.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC093089; AAS72371.1; -
DR InterPro: IPR004883; DUF260.
DR Pfam: PF03195; DUF260; 1.
DR PROSITE; PS50891; LOB; 1.
SQ SEQUENCE 145 AA; 14935 MW; 93868BA9B984502D CRC64;

Query Match 1.3%; Score 8; DB 2; Length 145;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 GGAASPPA 164
DB 108 GGAASPPA 115

RESULT 61
ID Q9NMA0 PRELIMINARY; PRT; 146 AA.
AC Q9NMA0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein FLJ10193.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN
RP SEQUENCE FROM N.A.
RA [1]
RA Tissue=Whole embryo;

```

```

RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niimura K., Itoh T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimura M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
RA Oono Y., Takiguchi S., Watanabe S., Yosida M., Houchi T., Kusano Y.,
RA Kanehori K., Takahashi-Fujii A., Hara R., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Miasahiro K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shibata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Taehiro H., Tanigami A., Fujiwara T.,
RA Oono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohnori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senda T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamaehita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RA "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK001055; BAA91483.1; -
DR InterPro: IPR011425; CSE2.
DR Pfam; PF07544; CSE2; 1.
SQ SEQUENCE 146 AA; 16403 MW; 24CDB7CBDFD36D1A CRC64;

Query Match 1.3%; Score 8; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 EQQQQQQ 241
DB 112 EQQQQQQ 119

RESULT 62
ID Q8WNG2 PRELIMINARY; PRT; 156 AA.
AC Q8WNG2;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Mi-ER1 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN
RP SEQUENCE FROM N.A.
RA [1]
RA Tissue=Lung;
RC MEDLINE=22368257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hale J.F.,
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uudin T.B., Toshivsky S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.V., Hultk S.W.,

```

```

RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017423; AAH17423.1; -.
DR Pfam; PF01448; ELM2; 1.
SQ SEQUENCE 156 AA; 17773 MW; 0075EB9819C4285A CRC64;

Query Match 1.3%; Score 8; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 416 FSSSEIDL 423
Db 8 FSSSEIDL 15

RESULT 63
O6ETP9 PRELIMINARY; PRT; 156 AA.
AC O6ETP9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein OJ1298_H07.10.
GN Name=OJ1298_H07.10;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxId=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004847; BAD28061.1; -.
KW Hypothetical protein.
SQ SEQUENCE 156 AA; 16409 MW; C4B43AD9AD7C4287 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 171 GPPPLPLP 178
Db 11 GPPPLPLP 18

RESULT 64
O6DGM8 PRELIMINARY; PRT; 156 AA.
AC O6DGM8;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Zgc:92851.
GN Name=zgc:92851;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxId=955;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallory S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
CC -1-SIMILARITY: Belongs to the bzip family.
DR EMBL; BC076313; AAH76313.1; -.
DR GO; GO:0005634; Cnucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR008917; Bk_transcr_DNA.
DR InterPro; IPR000837; LeuZip_Fos.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00170; bZIP_1; I.
DR PRINTS; PR00042; LEUZIPPFOS.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS02017; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 156 AA; 17914 MW; 0DE59F0B9676471D CRC64;

Query Match 1.3%; Score 8; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 247 IEELKSER 254
Db 116 IEELKSER 123

RESULT 65
O8SQ36 PRELIMINARY; PRT; 160 AA.
AC O8SQ36;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE NucIn (Fragment).
GN Name=MUC1/VNTR3;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxId=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=holslein;
RA Yamamoto N.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

```

DR EMBL: AB084275; BAB91141.1; -.
 FT NON_TER 1
 FT NON_TER 160
 SQ SEQUENCE 160 AA; 14163 MW; B709943E9269185F CRC64;

Query Match
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 GSPAPSPA 72
 DB 36 GSPAPSPA 43

RESULT 66
 Q8SQ37

ID Q8SQ37 PRELIMINARY; PRT; 160 AA.

AC Q8SQ37; 01-JUN-2002 (TEMBLrel. 21, Created)

DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)

DE 01-MAR-2004 (TEMBLrel. 26, Last annotation update)

Mucin (Fragment).

GN Name=MUC1/VNTR2;

OS Bos taurus (Bovine). Chordata; Craniata; Vertebrata; Euteleostomi;

OC Eukaryota; Metazoa; Chordata; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovinae; Bos.

NCBI_TaxID=9913;

RP [1] SEQUENCE FROM N.A.

RC STRAIN=holslein;

RA Yamamoto N.;

RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB084274; BAB91140.1; -.
 FT NON_TER 1
 FT NON_TER 160
 SQ SEQUENCE 160 AA; 14178 MW; 61CC8305BC403B03 CRC64;

Query Match
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 GSPAPSPA 72
 DB 96 GSPAPSPA 103

RESULT 67
 Q8SQ38

ID Q8SQ38 PRELIMINARY; PRT; 160 AA.

AC Q8SQ38; 01-JUN-2002 (TEMBLrel. 21, Created)

DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)

DE 01-MAR-2004 (TEMBLrel. 26, Last annotation update)

Mucin (Fragment).

GN Name=MUC1/VNTR1;

OS Bos taurus (Bovine). Chordata; Craniata; Vertebrata; Euteleostomi;

OC Eukaryota; Metazoa; Chordata; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovinae; Bos.

NCBI_TaxID=9913;

RP [1] SEQUENCE FROM N.A.

RC STRAIN=holslein;

RA Yamamoto N.;

RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB084273; BAB91139.1; -.
 FT NON_TER 1
 FT NON_TER 160
 SQ SEQUENCE 160 AA; 14194 MW; FC0CD5362DA90280 CRC64;

Query Match
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 65 GSPAPSPA 72
 DB 36 GSPAPSPA 43

RESULT 68
 Q8LN77

ID Q8LN77 PRELIMINARY; PRT; 162 AA.

AC Q8LN77; 01-OCT-2002 (TEMBLrel. 22, Created)

DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)

DE 05-JUL-2004 (TEMBLrel. 27, Last annotation update)

Hypothetical protein OSJNB0091N21.43.

GN ORFNames=OSJNB0091N21.43;

OS Oryza sativa (japonica cultivar-group).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzae; Oryza.

NCBI_TaxID=39947;

RP [1] SEQUENCE FROM N.A.

RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Kim M.M.,

RA Overton II L.L., Bera J.D., Teltrin T., Krol M.I., Jatrani B.B.,

RA Jin S.S., Koo H., Zismann V., Hsiao J., Blunt S., Vanaken S.S.,

RA Uteback T.T., Feldblyum T.V., Yang Q.Q., Haas B.J., Suh B.B.,

RA Peterson J.J., Quackenbush J., White O., Salzberg S.L., Fraser C.M.,

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RP [2] SEQUENCE FROM N.A.

RA Buell R.;

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

RP [3] SEQUENCE FROM N.A.

RA The Rice Chromosome 10 Sequencing Consortium;

RT "in-depth view of structure, activity, and evolution of rice

chromosome 10."

RL Science 300:1566-1569(2003).

RP [4] SEQUENCE FROM N.A.

RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;

RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC091122; AAM94929.1; -.
 DR EMBL: AE017111; AAP54554.1; -.
 DR Gramene; Q8LN77; -.
 KM Hypothetical protein.
 SQ SEQUENCE 162 AA; 16961 MW; 5E6C55612BCB9704 CRC64;

QY 10 AAAAPAGG 17
 DB 64 AAAAPAGG 71

RESULT 69
 Q9RBW6

ID Q9RBW6 PRELIMINARY; PRT; 162 AA.

AC Q9RBW6; 01-MAY-2000 (TEMBLrel. 13, Created)

DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)

DE Putative transposase.

OS Pseudomonas syringae (pv. phaseolicola).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

NCBI_TaxID=319;

RP [1] SEQUENCE FROM N.A.

RC STRAIN=Race 7;
RA MEDLINE=99415592; PubMed=10485919; DOI=10.1073/pnas.96.19.10875;
RA Jackson R.W., Athanassopoulos E., Tsimis G., Mansfield J.W.,
RA Sesma A., Arnold D.L., Gibbon M.J., Murrillo J., Taylor J.D.,
RA Vivian A.;
RT "Identification of a pathogenicity island, which contains genes for
RT virulence and avirulence, on a large native plasmid in the bean
RT pathogen *Pseudomonas syringae* pathovar *phaseolicola*.";
RT Proc. Natl. Acad. Sci. U.S.A. 96:10875-10880(1999).
RL EMBL; AF141883; AAD47208.1; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004803; F:transposase activity; IEA.
DR GO; GO:0006313; P:DNA transposition; IEA.
DR InterPro; IPR002513; Transposase_7.
DR Pfam; PF01526; Transposase_7; 2.
KW plasmid.
SQ SEQUENCE 162 AA; 17600 MW; A8BC8C149D191724 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 252 SERDITLA 259
Db 10 SERDITLA 17

RESULT 70
ID 06CE45 PRELIMINARY; PRT; 163 AA.
AC 06CE45;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DE Yarrowia lipolytica (TREMBlrel. 28, Last annotation update)
DE Yarrowia lipolytica chromosome B of strain CUB99 of Yarrowia
DE lipolytica.
GN ORFNames=YALI0B18656g;
OS Yarrowia lipolytica CUB99.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxId=284591;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CUB99;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangoul L., Aigle M., Anthonard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boistame A., Boyer J., Cattolico L., Confanioli F., de Darvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantreay F., Hennequin C., Jaumaux N., Joyet P., Kachouri R.,
RA Kerrest A., Kozul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Niclaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swemene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Weyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudon B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in Yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CUB99;
RG Genoscope;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382128; CAG83320.1; -.
SQ SEQUENCE 163 AA; 18631 MW; 2649A2C5E71B0A72 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 222 LKQITLLQ 229
Db 13 LKQITLLQ 20

RESULT 71
ID 073L27 PRELIMINARY; PRT; 165 AA.
AC 073L27;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE UVR domain protein.
DE UVR domain protein.
GN OrderedLocNames=TDE2038;
OS Treponema denticola.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxId=158;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35405 / DSM 14222;
RX PubMed=15064399; DOI=10.1073/pnas.0307639101;
RA Seshadri R., Myers G.S.A., Tettelin H., Eisen J.A., Heidelberg J.F.,
RA Dodson R.J., Davidson T.M., Deboy R.T., Fouts D.E., Haft D.H.,
RA Selenog J., Ren O., Brinkac L.M., Madupu R., Kolonay J.F.,
RA Durkin S.A., Daugherty S.C., Shetty J., Shvartsbeyn A.,
RA Gebregeorgis E., Geer K., Tegge G., Malek U.A., Ayodeji B.,
RA Shatman S., McLeod W.P., Smajic D., Howell J.K., Pal S., Amin A.,
RA Washist P., McNeill T.Z., Xiang Q., Sodergren E., Baca E.,
RA Weinstein G.M., Norris S.J., Fraser C.M., Paulsen I.T.;
RT "Comparison of the genome of the oral pathogen *Treponema denticola*
RT with other spirochete genomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:5646-5651(2004).
DR EMBL; AE017252; AAS12552.1; -.
DR TIGR; TDE2038; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004518; F:nuclease activity; IEA.
DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
DR InterPro; IPR001943; UVRB/C.
DR InterPro; IPR009055; UVRB_C.
DR Pfam; PF02151; UVR; 1.
DR PROSITE; PS50151; UVR; 1.
KW complete proteome.
SQ SEQUENCE 165 AA; 18565 MW; 75079265D3532539 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 165;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 DYERAAAL 31
Db 143 DYERAAAL 150

RESULT 72
ID 066EP9 PRELIMINARY; PRT; 172 AA.
AC 066EP9;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DE Hypothetical protein.
DE Hypothetical protein.
GN ORFNames=YPTB0643;
OS Yersinia pseudotuberculosis IP 32953.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxId=273123;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IP 32953;
RX PubMed=15358858;
RA Chain P.S.G., Cantiel E., Larimer F.W., Lamerdin J., Skoutland P.O.,
RA Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin L.V.,
RA Brubaker R.R., Fowler J., Himebusch B.J., Narceau W., Medigue C.,

RA Simonet M., Chenal-Francisque V., Souza B., Dacheux D., Elliott J.M.,
RA Derbise A., Hauser L.J., Garcia E.;
RT "Insights into the genome evolution of *Yersinia pestis* through whole
RT genome comparison with *Yersinia pseudotuberculosis*."
RL Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831(2004).
DR EMBL: BX36398; CAH19883.1; -.
DR InterPro: IPR008514; DUF796.
DR Pfam: PF05638; DUF796; 1.
KM Hypothetical protein.
SQ SEQUENCE 172 AA; 18841 MW; 7D6915B9DB0B1AF CRC64;

Query Match 1.3%; Score 8; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 LGGGGSGS 214
DB 44 LGGGGSGS 51

RESULT 73
08ZIJ6 PRELIMINARY; PRT; 172 AA.
ID 08ZIJ6 PRELIMINARY; PRT; 172 AA.
AC 08ZIJ6, Q74035; Q7CG93.
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein YPO0502.
GN Ordered locus names: YP0502, Y3673;
OS *Yersinia pestis*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Yersinia*.
OX NCBI_TaxID=632;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;
RA Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C.M., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Terriza A.-M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moulie S., Oyston P.C.F., Quail M.A., Rutherford K.M.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barril B.G.;
RT "Genome sequence of *Yersinia pestis*, the causative agent of plague."
RT Nature 413:523-527(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RX DOI=10.1126/jb.184.16.4601-4611.2002;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Featherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of *Yersinia pestis* KIM."
RT J. Bacteriol. 184:4601-4611(2002).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=91001 / Biovar Mediaevalis;
RX Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,
RA Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,
RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang F.,
RA Wang R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ14143; CAC89359.1; -.
DR EMBL: AE013970; AAM87221.1; -.
DR EMBL: AE017141; AAS63825.1; -.
DR PIR: AD0062; AD0062.
DR GO: GO:0005786; C:signal recognition particle (sensu Eukaryota); IEA.
DR GO: GO:0008312; F:7S RNA binding; IEA.
DR GO: GO:000605; P:protein targeting; IEA.

DR Pfam: PF05638; DUF796; 1.
KM Complete proteome; Hypothetical protein.
SQ SEQUENCE 172 AA; 18841 MW; 7D6915B9DB0B1AF CRC64;

Query Match 1.3%; Score 8; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 LGGGGSGS 214
DB 44 LGGGGSGS 51

RESULT 74
06UIP1 PRELIMINARY; PRT; 173 AA.
ID 06UIP1 PRELIMINARY; PRT; 173 AA.
AC 06UIP1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Heat shock protein 2 (Fragment).
GN Name=HSP2;
OS *Macaca mulatta* (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; *Macaca*.
OX NCBI_TaxID=9544;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22946167; PubMed=14557539; DOI=10.1073/pnas.2135499100;
RA Caceres M., Lachner J., Zapala M.A., Redmond J.C., Kudo L.,
RA Geschwind D.H., Lockhart D.J., Preuss T.M., Barlow C.;
RT "Elevated gene expression levels distinguish human from non-human
RT primate brains."
RT Proc. Natl. Acad. Sci. U.S.A. 100:13030-13035(2003).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
DR EMBL: AY369825; ARI1254.1; -.
DR GO: GO:0006457; P:protein folding; IEA.
DR GO: GO:0006986; P:response to unfolded protein; IEA.
DR InterPro: IPR01023; Hsp70.
DR Pfam: PF00012; Hsp70; 1.
KM ATP-binding; Heat shock.
FT NON TER 1
SQ SEQUENCE 173 AA; 19094 MW; 2117E01D3A89EB26 CRC64;

Query Match 1.3%; Score 8; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 209 GGGGSGGS 216
DB 157 GGGGSGGS 164

RESULT 75
06UIP2 PRELIMINARY; PRT; 174 AA.
ID 06UIP2 PRELIMINARY; PRT; 174 AA.
AC 06UIP2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Heat shock protein 2 (Fragment).
GN Name=HSP2;
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
OX NCBI_TaxID=9598;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22946167; PubMed=14557539; DOI=10.1073/pnas.2135499100;
RA Caceres M., Lachner J., Zapala M.A., Redmond J.C., Kudo L.,
RA Geschwind D.H., Lockhart D.J., Preuss T.M., Barlow C.;
RT "Elevated gene expression levels distinguish human from non-human

RT Primate brains.;
RL Proc. Natl. Acad. Sci. U.S.A. 100:13030-13035(2003).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
DR EMBL; AY369824; AAR1253.1; -;
DR GO; GO:0006457; P:protein folding; IEA.
DR GO; GO:0006986; P:response to unfolded protein; IEA.
DR InterPro: IPR01023; Hsp70.
DR Pfam; PF00012; Hsp70; 1.
KW ATP-binding; Heat shock.
FT NON_TER 1
SQ SEQUENCE 174 AA; 19151 MW; 4252FCF7593EB34 CRC64;

QY 209 GGGSGSGAS 216
 |||||
Db 158 GGGSGSGAS 165

Query Match 1.3%; Score 8; DB 2; Length 174;
Best Local Similarity 100.0%; Pred.No.1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Search completed: March 23, 2005, 15:52:05
Job time : 195 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2005, 15:42:59 ; Search time 42 Seconds
(without alignments)
1406.597 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 614

Sequence: 1 MTRSAVFKAAPAGGNPE.....RSRCRLRIQKKQPHRTCRK 614

Scoring table: OLIGO

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

PIR_79:*\n1: p1r1:*\n2: p1r2:*\n3: p1r3:*\n4: p1r4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	1.6	1289	184505	calcium-dependent
2	9	1.5	102	184505	hypothetical prote
3	9	1.5	334	602409	protein kinase C-b
4	9	1.5	577	109024	proline-rich prote
5	8	1.3	109	102039	acidic ribosomal p
6	8	1.3	109	102716	acidic ribosomal p
7	8	1.3	112	110267	ribosomal protein
8	8	1.3	114	110267	acidic ribosomal p
9	8	1.3	120	110267	hypothetical prote
10	8	1.3	136	110267	probable gln prot
11	8	1.3	172	110267	conserved hypotet
12	8	1.3	263	110267	probable oxidoredu
13	8	1.3	274	110267	homeobox protein H
14	8	1.3	274	110267	homeobox protein H
15	8	1.3	274	110267	homeobox protein H
16	8	1.3	289	110267	probable RNA-bind
17	8	1.3	291	110267	glycine-rich prote
18	8	1.3	330	110267	transcription fact
19	8	1.3	344	110267	oncogene homolog x
20	8	1.3	371	110267	alpha-helical coil
21	8	1.3	384	110267	PS14.12 protein
22	8	1.3	384	110267	hypothetical prote
23	8	1.3	392	110267	farneylelated prote
24	8	1.3	395	110267	CCAR/enhancer bin
25	8	1.3	410	110267	triglycer factor 2 l
26	8	1.3	418	110267	hypothetical prote
27	8	1.3	419	110267	hypothetical prote
28	8	1.3	444	110267	triglycer factor 2 l
29	8	1.3	449	110267	interferon-related

30	8	1.3	449	2	A30303	interferon-related
31	8	1.3	455	1	S01820	glycine-rich cell
32	8	1.3	469	2	S62702	H+-transporting tw
33	8	1.3	486	1	A57601	transcription fact
34	8	1.3	491	2	S17916	H+-transporting tw
35	8	1.3	506	2	S33922	H+-transporting tw
36	8	1.3	506	2	S46508	H+-transporting tw
37	8	1.3	507	1	PMRPA	H+-transporting tw
38	8	1.3	507	2	S12309	H+-transporting tw
39	8	1.3	507	2	JN0769	H+-transporting tw
40	8	1.3	507	2	A26760	H+-transporting tw
41	8	1.3	508	1	PMZAM	H+-transporting tw
42	8	1.3	508	2	S29792	H+-transporting tw
43	8	1.3	508	2	S26979	H+-transporting tw
44	8	1.3	509	1	PMNTAC	H+-transporting tw
45	8	1.3	509	1	PMRTAM	H+-transporting tw
46	8	1.3	509	1	PMRTAM	H+-transporting tw
47	8	1.3	509	2	T11937	H+-transporting tw
48	8	1.3	509	2	H97673	ATP synthase alpha
49	8	1.3	509	2	AG2898	H+-transporting tw
50	8	1.3	509	2	AD3283	H+-transporting tw
51	8	1.3	510	2	S19261	H+-transporting tw
52	8	1.3	510	2	S10997	H+-transporting tw
53	8	1.3	511	2	S07316	H+-transporting tw
54	8	1.3	512	2	S04672	H+-transporting tw
55	8	1.3	513	2	S25955	H+-transporting tw
56	8	1.3	546	2	T14604	hypothetical 59.6K
57	8	1.3	581	2	T22341	hypothetical prote
58	8	1.3	590	2	A26638	homeotic protein D
59	8	1.3	620	2	A53731	translation initia
60	8	1.3	633	2	A36353	DNA repair protein
61	8	1.3	639	2	A55719	dnak-type molecula
62	8	1.3	689	2	A43800	nuclear autoantige
63	8	1.3	696	2	T01209	search synthase (E
64	8	1.3	737	2	S28030	DNA-binding protei
65	8	1.3	748	2	S61247	DNA helicase/prima
66	8	1.3	785	2	S54016	SOX2 protein - yea
67	8	1.3	873	2	B41054	fasciclin II PI-11
68	8	1.3	876	2	T49801	hypothetical prote
69	8	1.3	881	2	T01269	serine/threonine-s
70	8	1.3	1051	2	S55259	TIF1 protein - mou
71	8	1.3	1099	2	A56155	tumor suppressor p
72	8	1.3	1259	2	S25954	gene alpha introm 2
73	8	1.3	1366	2	B86292	PH2.12 protein -
74	8	1.3	1395	2	S25997	gene alpha introm 1
75	8	1.3	1597	2	T08428	gene small optic 1
76	8	1.3	1606	2	T49219	translation initia
77	8	1.3	1957	2	T38077	hypothetical colle
78	8	1.3	2475	2	T00047	gelatin lyase (EC 4
79	8	1.3	4957	2	T03455	ALR protein - huma
80	8	1.3	5262	2	T03454	ALR protein - huma
81	8	1.3	26	2	C44379	omega-conotoxin SV
82	8	1.3	26	2	A32806	transposase - Ech
83	8	1.3	36	1	A48850	chloride channel 1
84	8	1.3	44	4	S43173	probable serine pr
85	8	1.3	63	2	S24422	4-oxalocrotonate t
86	8	1.3	64	2	A86333	hypothetical prote
87	8	1.3	70	2	F69818	hypothetical prote
88	8	1.3	80	2	T11050	phosphotransferase
89	8	1.3	85	1	I64137	hypothetical prote
90	8	1.3	86	2	G81950	hypothetical prote
91	8	1.3	88	2	E90169	hypothetical prote
92	8	1.3	94	2	F31844	spda protein - Str
93	8	1.3	97	2	G60110	repetitive protein
94	8	1.3	97	2	E64509	hypothetical prote
95	8	1.3	104	2	G81153	hypothetical prote
96	8	1.3	104	2	H83876	hypothetical prote
97	8	1.3	104	2	T02612	hypothetical prote
98	8	1.3	117	1	L2MSME	Ig lambda-2 chain
99	8	1.3	118	2	T38835	probable 60s ribos
100	8	1.3	122	2	D75323	ribosomal protein

ALIGNMENTS

RESULT 1
184505
A:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 09-Jul-2004
C:Accession: 184505
R:Walent, J.H.; Porter, B.W.; Martin, T.F.
Cell 70, 765-775, 1992
A>Title: A novel 145 kd brain cytosolic protein reconstitutes Ca(2+)-regulated secretion
A:Reference number: 139188; MUID:92386596; PMID:1516133
A:Accession: 184505
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1289 <RES>
A:Cross-references: UNIPROT:Q62717; EMBL:U16802; NID:9577427; PIDN:AAB8635.1; PID:95774
C:Genetics:
A:Gene: CAPS
C:Superfamily: calcium-dependent actin-binding protein; pleckstrin repeat homology
C:Keywords: actin binding

Query Match 1.6%; Score 10; DB 2; Length 1289;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGGAS 217
|||||
Db 60 GGGGGSGGAS 69

RESULT 2
F64341
A:hypothetical protein MJ034 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: F64341
R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
J.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
reson, J.D.; Sadow, P.W.; Hanna, M.C.; Coton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Frazer, C.M.; Smith, H.O.; Moese, C
A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: F64341
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-102 <BUL>
A:Cross-references: UNIPROT:Q57780; GB:U67487; GB:L77117; NID:91591043; PIDN:AAB98322.1;
C:Genetics:
A:Map position: FOR309807-310115
A:Start codon: GTG

Query Match 1.5%; Score 9; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 KEKEIEELK 251
|||||
Db 72 KEKEIEELK 80

RESULT 3
G02409
A:protein kinase C-binding protein RACK17 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C:Accession: G02409
R:Kuroda, S.; Tokunaga, C.; Kiyohara, Y.; Konishi, H.; Kikkawa, U.
submitted to the EMBL Data Library, February 1996
A:Reference number: H01212
A:Accession: G02409

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-334 <KUR>
A:Cross-references: EMBL:U48250; NID:91199656; PIDN:AAC72247.1; PID:91199657

Query Match 1.5%; Score 9; DB 2; Length 334;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGGAS 216
|||||
Db 285 GGGGGSGGAS 293

RESULT 4
T09024
A:proline-rich protein T27E11.90 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T09024
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T09024
A:Molecule type: DNA
A:Residues: 1-577 <BEV>
A:Cross-references: UNIPROT:Q9STP1; EMBL:AL078579; GSPDB:GN00062; ATSP:T27E11.90
A:Experimental source: cultivar Columbia; BAC clone T27E11
C:Genetics:
A:Gene: ATSP:T27E11.90
A:Map position: 4
A:Introns: 26/1, 117/1, 338/1, 411/3, 430/2, 498/2

Query Match 1.5%; Score 9; DB 2; Length 577;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 PLPLPGPPP 182
|||||
Db 198 PLPLPGPPP 206

RESULT 5
T02039
A:acidic ribosomal protein Pia - maize
C:Species: Zea mays (maize)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 09-Jul-2004
C:Accession: T02039
R:Bailey-Serres, U.; Vangala, S.; Szick, K.; Lee, C.H.
Plant Physiol. 114, 1293-1305, 1997
A>Title: Acidic phosphoprotein complex of the 60S ribosomal subunit of maize seedling ro
A:Reference number: Z14507; MUID:97422884; PMID:9276949
A:Accession: T02039
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-109 <BAI>
A:Cross-references: UNIPROT:P52855; EMBL:U62752; NID:92431768; PIDN:AAB71079.1; PID:92431
A:Experimental source: strain B73
C:Genetics:

A:Gene: rpp1a
C:Superfamily: rat acidic ribosomal protein P1
C:Keywords: protein biosynthesis; ribosome

Query Match 1.3%; Score 8; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAGG 17
|||||
Db 72 AAAAPAGG 79

RESULT 6

T02716
acidic ribosomal protein P1 - maize
N:Alternate names: ribosomal protein L12
C/Species: Zea mays (maize)
C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C/Accession: T02716
R/Hamilton, D.A.; Turcich, M.P.; Bokhari-Riza, A.; Mascarenhas, J.P.
submitted to the EMBL Data Library, November 1995
A/Reference number: 214703
A/Accession: T02716
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-109 <HMM>
A/Cross-references: UNIPROT:P52855; EMBL:U40147; NID:G1209700; PIDN:AAA91168.1; PID:G120
A/Experimental source: strain W-22
C/Superfamily: rat acidic ribosomal protein P1
C/Keywords: protein biosynthesis; ribosome

Query Match 1.3%; Score 8; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17
Db 72 AAAAPAG 79

RESULT 7
T10267
ribosomal protein L21 - *Osccheius brevesophaga*
C/Species: *Osccheius brevesophaga*
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C/Accession: T10267
R/Evans, D.; Zorio, D.; MacMorris, M.; Winter, C.E.; Lea, K.; Blumenthal, T.
Proc. Natl. Acad. Sci. U.S.A. 94, 9751-9756, 1997
A/Title: Operons and SL2 trans-splicing exist in nematodes outside the genus *Caenorhabditis*
A/Reference number: 217008; MUID:97420754; PMID:9275196
A/Accession: T10267
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-112 <EVA>
A/Cross-references: UNIPROT:O01359; EMBL:U90830; NID:G1906641; PID:G1906643
A/Experimental source: strain CEM1
C/Genetics:
A/Gene: rpl-21
A/Intons: 20/3; 57/3
C/Superfamily: rat acidic ribosomal protein P1
C/Keywords: ribosome

Query Match 1.3%; Score 8; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17
Db 75 AAAAPAG 82

RESULT 8
R5CH2E
acidic ribosomal protein P1 - chicken
N:Alternate names: ribosomal phosphoprotein P1; ribosomal protein CU12ELI
C/Species: *Gallus gallus* (chicken)
C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 09-Jul-2004
C/Accession: S02029
R/Ferrio, J.A.; Reznach, F.C.
Eur. J. Biochem. 177, 513-516, 1988
A/Title: The complete sequence of a chicken-muscle cDNA encoding the acidic ribosomal p
A/Reference number: S02029; MUID:89064804; PMID:3197716
A/Accession: S02029
A/Molecule type: mRNA
A/Residues: 1-114 <FER>
A/Cross-references: UNIPROT:P16660; EMBL:X13876; NID:G63063; PIDN:CAA32080.1; PID:G63064

A/Note: the sequence from Fig. 4 is inconsistent with that from Fig. 1 in lacking 96-Glu
C/Superfamily: rat acidic ribosomal protein P1
C/Keywords: phosphoprotein; protein biosynthesis; ribosome

Query Match 1.3%; Score 8; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17
Db 71 AAAAPAG 78

RESULT 9
H95330
hypothetical protein Sma1018 [imported] - *Sinorhizobium meliloti* (strain 1021) magaplaasm
C/Species: *Sinorhizobium meliloti*
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C/Accession: H95330
R/Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
-; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A/Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*
A/Reference number: A95262; MUID:21396509; PMID:11481432
A/Accession: H95330
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-120 <KIR>
A/Cross-references: UNIPROT:Q922D6; GB:AE006469; PIDN:AAK65210.1; PID:G14523657; GSPDB:GI
A/Experimental source: strain 1021, megaplaasmid pSymA
R/Galibert, F.; Finan, T.M.; Long, S.R.; Pulver, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A/Title: The complete genome of the legume symbiont *Sinorhizobium meliloti*.
A/Reference number: A96039; MUID:21368234; PMID:11474104
A/Contents: annotation
C/Genetics:
A/Gene: Sma1018
A/Genome: plasmid

Query Match 1.3%; Score 8; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 LLLQLDLI 233
Db 104 LLLQLDLI 111

RESULT 10
C70761
probable glbN protein - *Mycobacterium tuberculosis* (strain H37RV)
C/Species: *Mycobacterium tuberculosis*
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C/Accession: C70761
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: C70761
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-136 <COL>
A/Cross-references: UNIPROT:Q10784; GB:Z74020; GB:AL123456; NID:G3261584; PIDN:CAA98320.1
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: glbN

C/Superfamily: Tetrahymena globin

Query Match 1.3%; Score 8; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 AAALGGPE 35
DB 63 AAALGGPE 70

RESULT 11

AD0062 conserved hypothetical protein YPO0502 [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C/Accession: AD0062
R/Parhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, N.
A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A/Reference number: AB0001; MUID:21470413; PMID:11586360
A/Accession: AD0062

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-172 <RUR>
A/Cross-references: UNIPROT:Q8ZIU6; GB:AL590842; PIDN:CAC8935.1; PID:G15978596; GSPDB:G
C/Genetics:
A/Gene: YPO0502

Query Match 1.3%; Score 8; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LGGGGSG 214
DB 44 LGGGGSG 51

RESULT 12

T50596 Probable oxidoreductase [imported] - Streptomyces coelicolor

C/Species: Streptomyces coelicolor
C/Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C/Accession: T50596
R/Redenbach, M.; Kisser, H.M.; Denaplatte, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopy
Mol. Microbiol. 21, 77-96, 1996
A/Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S
A/Reference number: Z20556; MUID:97000351; PMID:8843436
A/Accession: T50596
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-263 <RED>
A/Cross-references: UNIPROT:Q9RKP1; EMBL:AL133220; PIDN:CAB61733.1
A/Experimental source: strain A3 (2)
C/Genetics:
A/Note: SCC75A.33

C/Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 1.3%; Score 8; DB 2; Length 263;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAGG 17
DB 56 AAAAPAGG 63

RESULT 13

A84617 homeobox domain transcription factor (HAT9) [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004

C/Accession: A84617
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.R.; Umayam, L.; Tallon, L.;
eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:1067197
A/Accession: A84617
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-274 <STO>
A/Cross-references: UNIPROT:P46603; GB:AE002093; NID:G3445197; PIDN:AAC32427.1; GSPDB:GN

C/Genetics:
A/Gene: ATG22800
A/Map position: 2

C/Superfamily: unassigned homeobox proteins; homeobox homology
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 1.3%; Score 8; DB 2; Length 274;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGA 215
DB 234 GGGGSGGA 241

RESULT 14

T52371 homeobox protein HAT9 [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 16-Aug-2004
C/Accession: T52371
R/Schena, M.; Davis, R.W.
Proc. Natl. Acad. Sci. U.S.A. 91, 8393-8397, 1994
A/Title: Structure of homeobox-leucine zipper genes suggests a model for the evolution of
A/Reference number: Z11420
A/Accession: T52371
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-274 <SCH>
A/Cross-references: UNIPROT:P46603; EMBL:U09341; PIDN:AAA56907.1
C/Genetics:
A/Gene: HAT9

C/Superfamily: homeobox homology
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 1.3%; Score 8; DB 2; Length 274;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGA 215
DB 234 GGGGSGGA 241

RESULT 15

T52372 homeobox protein HAT9 [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 16-Aug-2004
C/Accession: T52372
R/Schena, M.; Davis, R.W.
Proc. Natl. Acad. Sci. U.S.A. 91, 8393-8397, 1994
A/Title: Structure of homeobox-leucine zipper genes suggests a model for the evolution of
A/Reference number: Z11420
A/Accession: T52372
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-274 <SCH>
A/Cross-references: UNIPROT:P46603; EMBL:U09342; PIDN:AAA56908.1
C/Genetics:

A:Gene: HAT9
 A:introns: 137/3, 164/2
 C:Superfamily: homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 1.3%; Score 8; DB 2; Length 274;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
 Db 234 GGGGGSGA 241

RESULT 16

A84790
 probable RNA-binding protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
 C/Accession: A84790
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon, L.; Euse, D.; Niemman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A/Reference number: A84420; MUID:20083487; PMID:10617197
 A/Accession: A84790
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-289 <STO>
 A/Cross-references: UNIPROT:Q9ZU04; GB:A802093; NID:g4056477; PIDN:AAC98043.1; GSPDB:GN
 C/Genetics:
 A:Gene: At2g37220
 A/Map position: 2
 C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein

Query Match 1.3%; Score 8; DB 2; Length 289;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
 Db 193 GGGGGSGA 200

RESULT 17

S31415
 glycine-rich protein GRP22 - rape

C:Species: Brassica napus (rape)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C/Accession: S31415
 R:Bergeron, D.; Boivin, R.; Baezcyanski, C.L.; Bellemare, G.
 A:Submitted to the EMBL Data Library, August 1992
 A>Description: Characterization and expression of a gene family encoding glycine-rich pr
 A/Reference number: S31415
 A/Accession: S31415
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-291 <BBR>
 A/Cross-references: UNIPROT:Q39337; EMBL:Z15045; NID:g17820; PIDN:CAA78762.1; PID:g17821
 C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 1.3%; Score 8; DB 1; Length 291;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
 Db 217 GGGGGSGA 224

RESULT 18

T51834

transcription factor DREB2B, drought-induced [validated] - Arabidopsis thaliana

N/Alternate names: dehydration-responsive element-binding protein 2B
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004
 C/Accession: T51834
 R:Lin, X.; Kasuga, M.; Sakuma, Y.; Abe, H.; Miura, S.; Yamaguchi-Shinozaki, K.; Shinozaki, K.
 A>Title: Two transcription factors, DREB1 and DREB2, with an ERBBP/AP2 DNA binding domain, respectively, in Arabidopsis.
 A/Reference number: 225477
 A/Accession: T51834
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-330 <LIU>
 A/Cross-references: UNIPROT:O82133; EMBL:AB007791; PIDN:BA33795.1
 A/Experimental source: strain Colombia

A/Function:
 A>Description: binds to the DRE sequence in vitro and activates the transcription of the
 A/Note: expression induced by dehydration
 C/Keywords: transcription factor

Query Match 1.3%; Score 8; DB 2; Length 330;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 EQQQQQQLQ 241
 Db 242 EQQQQQQLQ 249

RESULT 19

A48101
 oncogene homolog xL-myc1 - African clawed frog

C:Species: Xenopus laevis (African clawed frog)
 C>Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C/Accession: A48101
 R:Schreiber-Agus, N.; Torres, R.; Horner, J.; Lau, A.; Jamrich, M.; DePinho, R.A.
 A>Title: Comparative analysis of the expression and oncogenic activities of Xenopus c-, l
 A/Reference number: A48101; MUID:93204991; PMID:8455622
 A/Accession: A48101
 A/Status: preliminary
 A/Molecule type: nucleic acid
 A/Residues: 1-344 <SCH>
 A/Cross-references: UNIPROT:Q05404; GB:L11362; NID:g214906; PIDN:AAA49994.1; PID:g214907
 A/Note: sequence extracted from NCBI backbone (NCBI:127681, NCBI:127683)
 C:Superfamily: myc transforming protein; myc transforming protein homology
 C/Keywords: DNA binding
 F:1-344/Domain: myc transforming protein homology <MYC>

Query Match 1.3%; Score 8; DB 2; Length 344;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 HSPKKEP 368
 Db 222 HSPKKEP 229

RESULT 20

A44122
 alpha-helical coiled coil protein Tlpa - Salmonella typhimurium

C:Species: Salmonella typhimurium
 C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C/Accession: A44122
 R:Koski, P.; Saarilahti, H.; Sukupolvi, S.; Taira, S.; Rikonen, P.; Osterlund, K.; Hurme, J.; Biol. Chem. 267, 12258-12265, 1992
 A>Title: A new alpha-helical coiled coil protein encoded by the salmonella typhimurium v
 A/Reference number: A44122; MUID:92291112; PMID:1601892
 A/Accession: A44122
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-371 <KOS>

A:Cross-references: UNIPROT:Q56080; GB:M88208; NID:g968910; PIDN:AAA74964.1; PID:g968911
A:Experimental source: large virulence-associated plasmid pEX102
A:Note: sequence extracted from NCBI backbone (NCBIN:106718, NCBI:P.106719)

Query Match 1.3%; Score 8; DB 2; Length 371;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 EELKSERD 255
|||||
DB 224 EELKSERD 231

RESULT 21

D86448
F5D14.12 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D86448
R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
C.M.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzalli,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86448

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-384 <STO>
A:Cross-references: UNIPROT:Q9LQW3; GB:AE05172; NID:g8920610; PIDN:AAF81332.1; GSPDB:GN
C:Genetics:
A:Map position: 1

Query Match 1.3%; Score 8; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LGGGGSGS 214
|||||
DB 241 LGGGGSGS 248

RESULT 22

G87332
hypothetical protein CC0674 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C:Accession: G87332

R:Hieman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hart, D.H.; Kolon,
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.;
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: G87332

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <STO>
A:Cross-references: UNIPROT:Q9NAC8; GB:AE05673; NID:g13421893; PIDN:AAK22659.1; GSPDB:G
C:Genetics:
A:Gene: CC0674

Query Match 1.3%; Score 8; DB 2; Length 389;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 72 FKAAAPAA 79

RESULT 23

T48360
farneoylated protein-like - Arabidopsis thaliana
N:Alternate names: protein F12E4.120
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T48360
R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.;
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24492

A:Accession: T48360
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-392 <BEV>
A:Cross-references: UNIPROT:Q9LZF1; EMBL:AL162751
A:Experimental source: cultivar Columbia; BAC clone F12E4
C:Genetics:
A:Map position: 5
A:Introns: 3/3; 50/1; 153/3; 180/1
A:Note: F12E4.120

Query Match 1.3%; Score 8; DB 2; Length 392;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAGG 17
|||||
DB 232 AAAAPAGG 239

RESULT 24

I49575
CCAAT/enhancer binding protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C:Accession: I49575
R:Christy, R.J.; Kaestner, K.H.; Gelman, D.E.; Lane, M.D.
Proc. Natl. Acad. Sci. U.S.A. 88, 2593-2597, 1991
A:Title: CCAAT/enhancer binding protein gene promoter: binding of nuclear factors during
A:Reference number: I49575; MUID:9112853; PMID:2006196
A:Accession: I49575

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-395 <RES>
A:Cross-references: UNIPROT:P53566; GB:M62362; NID:g192381; PIDN:AAA37374.1; PID:g192382
C:Genetics:
A:Gene: mc/EPP
C:Superfamily: CCAAT/enhancer-binding protein alpha

Query Match 1.3%; Score 8; DB 2; Length 395;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
|||||
DB 263 GGGGGSGA 270

RESULT 25

I40755
trigger factor 2 [validated] - Campylobacter jejuni
N:Contents: peptidylprolyl isomerase (EC 5.2.1.8), ribosome-bound
C:Species: Campylobacter jejuni
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: I40755
R:Griffiths, P.L.; Park, R.W.; Connerston, I.F.
Microbiology 141, 1359-1367, 1995
A:Title: The gene for Campylobacter trigger factor: evidence for multiple transcription
A:Reference number: I40754; MUID:95400450; PMID:7670637
A:Accession: I40755

A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-410 <RES>
A/Cross-references: UNIPROT:Q46108; EMBL:X85954; NID:G757795; PIDD:CAA59930.1; PID:G7577
A/Genetics: TG2
A/Accession: I40756
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 23-410 <RE2>
A/Cross-references: EMBL:X85954; NID:G757795; PIDD:CAA59931.1; PID:G757798
A/Genetics: TG3
A/Genetics: <TG2>
A/Genetics: tlgp2
A/Genetics: <TG3>
A/Genetics: tlgp3
C/Superfamily: FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor); BKBP-type
C/Keywords: cis-trans-isomerase
F/1-410/Product: trigger factor 2 #status predicted <TF2>
F/23-410/Product: trigger factor 3 #status predicted <TF3>
F/165-209/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>

Query Match 1.3%; Score 8; DB 1; Length 410;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 244 EKEIEELK 251
DB 331 EKEIEELK 338

RESULT 26
T15142
hypothetical protein T28F2.6 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C/Accession: T15142
R/Madsen, C.; Fronick, B.
submitted to the EMBL Data Library, April 1997
A/Description: The sequence of C. elegans cosmid T28F2.
A/Reference number: Z18300
A/Accession: T15142
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-418 <MAD>
A/Cross-references: UNIPROT:O01662; EMBL:AF000198; NID:G2047345; PID:G2047346; PIDD:AAB5
A/Experimental source: strain Bristol N2; clone T28F2
C/Genetics:
A/Genes: CESP:T28F2.6
A/Map position: 1
A/Intons: 49/3
C/Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 1.3%; Score 8; DB 2; Length 418;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAAPAG 17
DB 317 AAAAPAG 324

RESULT 27
G70602
hypothetical protein Rv1004c - Mycobacterium tuberculosis (strain H37Rv)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C/Accession: G70602
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634220
A/Accession: G70602
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-419 <COL>
A/Cross-references: UNIPROT:O05589; GB:294752; GB:AL123456; NID:G3261731; PIDD:CAB08158.1
A/Experimental source: strain H37Rv
C/Genetics:
A/Genes: Rv1004c

Query Match 1.3%; Score 8; DB 2; Length 419;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 176 PLRGPPL 183
DB 392 PLRGPPL 399

RESULT 28
H81437
trigger factor tlg Cj0193c [similarity] - Campylobacter jejuni (strain NCTC 11168)
N/Contains: peptidylprolyl isomerase (EC 5.2.1.8), ribosome-bound
C/Species: Campylobacter jejuni
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C/Accession: H81437
R/Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A/Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A/Reference number: A81250; MUID:20150912; PMID:10688204
A/Accession: H81437
A/Molecule type: DNA
A/Residues: 1-444 <PAR>
A/Cross-references: UNIPROT:Q46108; GB:AL139074; GB:AL111168; NID:G6967505; PIDD:CAB72676
A/Experimental source: serotype O2, strain NCTC 11168
C/Genetics:
A/Genes: tlg Cj0193c
C/Superfamily: FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor); BKBP-type
C/Keywords: cis-trans-isomerase
F/165-209/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>

Query Match 1.3%; Score 8; DB 2; Length 444;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 244 EKEIEELK 251
DB 331 EKEIEELK 338

RESULT 29
A44989
interferon-related protein T1S7 - mouse
N/Alternate names: IFN-beta-2; interferon beta-2
C/Species: Mus musculus (house mouse)
C/Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 09-Jul-2004
C/Accession: A44989; A05277
R/Varnum, B.C.; Lim, R.W.; Herchman, H.R.
Oncogene 4, 1263-1265, 1989
A/Title: Characterization of T1S7, a gene induced in Swiss 3T3 cells by the tumor promoter
A/Reference number: A44989; MUID:90016295; PMID:2797820
A/Accession: A44989
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-449 <VAR>
A/Cross-references: UNIPROT:P19182; GB:X17400; NID:G54805; PIDD:CAA5258.1; PID:G54806
R/Stup, D.; Windass, J.D.; Sor, F.; George, H.; Williams, B.R.G.; Fukuhara, H.; De Maeyer
Nucleic Acids Res. 10, 3069-3084, 1982
A/Title: Molecular cloning of partial cDNA copies of two distinct mouse IFN-beta mRNAs.
A/Reference number: A93424; MUID:82247191; PMID:6179042
A/Accession: A05277

A:Molecule type: mRNA
A:Residues: 386-449 <SKU>
C:Superfamily: Interferon alpha
C:Keywords: antiviral; transmembrane protein

Query Match 1.3%; Score 8; DB 2; Length 449;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 208 GGGGSGS 215
DB 15 GGGGSGS 22

RESULT 30
A30303
Interferon-related protein PC4 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: A30303
R:Illione, F.; Shooter, E.M.
Proc. Natl. Acad. Sci. U.S.A. 86, 2088-2092, 1989
A:Title: Early gene regulation by nerve growth factor in PC12 cells: induction of an int
A:Reference number: A30303; MUID:89184573; PMID:2467301
A:Accession: A30303
A:Molecule type: mRNA
A:Residues: 1-449 <TIR>
A:Cross-references: UNIPROT:P20695; GB:U04511; NID:g204972; PIDN:AAC28946.1; PID:g204974
C:Genetics:
A:Gene: PC4
C:Superfamily: Interferon alpha
C:Keywords: nucleus; transmembrane protein

Query Match 1.3%; Score 8; DB 2; Length 449;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 208 GGGGSGS 215
DB 15 GGGGSGS 22

RESULT 31
S01820
glycine-rich cell wall protein 1.8 precursor - kidney bean
C:Species: Phaseolus vulgaris (kidney bean)
C>Date: 30-Sep-1989 #sequence_revision 19-May-1994 #text_change 09-Jul-2004
C:Accession: S01820
R:Keller, B.; Sauer, N.; Lamb, C.J.
EMBO J. 7, 3625-3633, 1988
A:Title: Glycine-rich cell wall proteins in bean: gene structure and association of the
A:Reference number: S01820; MUID:89091109; PMID:3208742
A:Accession: S01820
A:Molecule type: DNA
A:Residues: 1-465 <KEL>
A:Cross-references: UNIPROT:P10496; EMBL:X13596; NID:g21002; PIDN:CAA1932.1; PID:g21003
C:Comment: This protein is enriched in the cell wall fraction of young hypocotyls and ov
tys.
C:Comment: Much of the sequence consists of tandemly repeated 22-residue segments with t
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8
C:Keywords: cell wall; structural protein; tandem repeat
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-465/Product: glycine-rich cell wall protein 1.8 #status predicted <MAT>

Query Match 1.3%; Score 8; DB 1; Length 465;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 208 GGGGSGS 215
DB 142 GGGGSGS 149

RESULT 32

S62702
H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - Platymonas subcordiformis n
C:Species: mitochondrion Platymonas subcordiformis
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C:Accession: S62702
R:Kessler, U.; Zetsche, K.
Plant Mol. Biol. 29, 1081-1086, 1995
A:Title: Physical map and gene organization of the mitochondrial genome from the unicell
A:Reference number: S62702; MUID:96145517; PMID:855450
A:Accession: S62702

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-469 <RES>
A:Cross-references: UNIPROT:Q36517; EMBL:Z47795; NID:g633584; PIDN:CAA87749.1; PID:g63358
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995
C:Genetics:
A:Gene: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxi
F:204-376/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 469;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 163 AATASDPA 170
DB 235 AATASDPA 242

RESULT 33
A57601
transcription factor dGATAC - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 08-Feb-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C:Accession: A57601
R:Lin, W.H.; Huang, L.H.; Yeh, J.Y.; Hoheisel, J.; Lehrach, H.; Sun, Y.H.; Tsai, S.F.
J. Biol. Chem. 270, 25150-25158, 1995
A:Title: Expression of a Drosophila GATA transcription factor in multiple tissues in the
n.
A:Reference number: A57601; MUID:96027621; PMID:7559649
A:Accession: A57601
A:Molecule type: mRNA
A:Residues: 1-486 <LIN>
A:Cross-references: UNIPROT:P91623; GB:D50542; NID:g1060873; PIDN:BAA09102.1; PID:g18156
C:Genetics:
A:Gene: FlyBase:Gata-C
A:Cross-references: FlyBase:Fgmn0001138
C:Superfamily: transcription factor dGATAC; GATA-type zinc finger homology
C:Keywords: DNA binding; transcription factor; zinc finger
F:258-311/Domain: GATA-type zinc finger homology <GZPF>
F:318-371/Domain: GATA-type zinc finger homology <GZPF>

Query Match 1.3%; Score 8; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 207 LGGGGSG 214
DB 72 LGGGGSG 79

RESULT 34

S17916
H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - evening primrose mitochondr
C:Species: mitochondrion Oenothera villaricae (evening primrose)
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C:Accession: S17916
R:Schuster, W.; Ternes, R.; Knopp, V.; Hiesel, R.; Wiesinger, B.; Brennicke, A.
Curr. Genet. 20, 397-404, 1991
A:Title: Distribution of RNA editing sites in Oenothera mitochondrial mRNAs and rRNAs.
A:Reference number: S17916; MUID:92224283; PMID:1725505

A:Accession: S17916
A:Molecule type: mRNA
A:Residues: 1-491 <SCH1>
A:Cross-references: UNIPROT:Q7M2G6
A:Accession: S20916
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-451, 'P', 453-476, 'P', 478-491 <SCH2>
C:Genetics:
A:Gene: atpA
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F;171-178/Region: nucleotide-binding motif A (P-loop)
F;205-357/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 491;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 AATASDPA 170
Db 236 AATASDPA 243

RESULT 35
S33922
H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - sugar beet mitochondrion
N:Alternate names: ATPase alpha chain
C:Species: mitochondrion Beta vulgaris var. altissima (sugar beet)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: S33922
R:Sendu, M.; Mikami, T.; Kinoshita, T.
Curr. Genet. 24, 164-170, 1993
A:Title: The sugar beet mitochondrial gene for the ATPase alpha-subunit: sequence, trans
A:Reference number: S33922; MUID:93365027; PMID:8358823
A:Accession: S33922
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-506 <SEN>
A:Cross-references: UNIPROT:Q06735; GB:D15065; NID:g285631; PIDN:BA03664.1; PID:g285632
C:Genetics:
A:Gene: atpA
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F;171-178/Region: nucleotide-binding motif A (P-loop)
F;205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 506;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 AATASDPA 170
Db 236 AATASDPA 243

RESULT 36
S46508
H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - beet mitochondrion
N:Alternate names: ATPase alpha chain
C:Species: mitochondrion Beta vulgaris (beet)
C:Date: 15-Jul-1995 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004
C:Accession: S46508; S46507; S51586; S29535; S29536
R:Xue, Y.; Collin, S.; Davies, D.R.; Thomas, C.M.
Plant Mol. Biol. 25, 91-103, 1994
A:Title: Differential screening of mitochondrial cDNA libraries from male-fertile and cy
A:Reference number: S46505; MUID:94272017; PMID:8003700
A:Accession: S46508
A:Molecule type: DNA
A:Residues: 1-506 <XUE>
A:Cross-references: UNIPROT:Q36862; EMBL:X68691

A:Experimental source: male-sterile CMS
A>Note: the authors translated the codon GAT for residue 207 as Val
A:Accession: S46507
A:Molecule type: DNA
A:Residues: 1-506 <XUF>
A:Cross-references: EMBL:X68690
A:Experimental source: male-fertile MF
A>Note: the authors translated the codon GAT for residue 207 as Val
R:Xue, Y.
submitted to the EMBL Data Library, October 1992
A:Reference number: S51586
A:Accession: S51586
A:Molecule type: DNA
A:Residues: 1-447, 'L', 449, 'G', 451-453, 'D', 455-506 <XUW>
A:Cross-references: EMBL:X68691; NID:g396759; PIDN:CAA48650.1; PID:g396760
A:Experimental source: male-sterile CMS
R:Xue, Y.; Davies, D.R.; Collin, S.; Thomas, C.M.
submitted to the EMBL Data Library, October 1992
A:Description: Molecular characterization of mitochondrial ATPase subunit A from cytoplas
A:Reference number: S29535
A:Accession: S29535
A:Molecule type: DNA
A:Residues: 1-447, 'L', 449, 'G', 451-453, 'D', 455-506 <XUN>
A:Cross-references: EMBL:X68690; NID:g11262; PIDN:CAA48649.1; PID:g11263
A:Experimental source: male-fertile MF
C:Genetics:
A:Gene: atpA
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F;171-178/Region: nucleotide-binding motif A (P-loop)
F;205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 506;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 AATASDPA 170
Db 236 AATASDPA 243

RESULT 37
PMRA
H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - rape mitochondrion
C:Species: mitochondrion Brassica napus (rape)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C:Accession: S13382; S30090
R:Handa, H.; Nakajima, K.
Plant Mol. Biol. 16, 361-364, 1991
A:Title: Nucleotide sequence and transcription analyses of the rapeseed (Brassica napus L
A:Reference number: S13382; MUID:91370881; PMID:1832575
A:Accession: S13382
A:Molecule type: DNA
A:Residues: 1-507 <HAN>
A:Cross-references: UNIPROT:P22201; EMBL:X56008; NID:g19553; PIDN:CAA39483.1; PID:g19554
A:Experimental source: strain Polima
R:Bonhomme, S.; Budar, F.; Lancelin, D.; Small, I.; Defrance, M.C.; Pelletier, G.
Mol. Gen. Genet. 235, 340-348, 1992
A:Title: Sequence and transcript analysis of the Nco2.5 Ogura-specific fragment correlate
A:Reference number: S30089; MUID:93101139; PMID:1281515
A:Accession: S30090
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-111 <BON>
A:Cross-references: EMBL:Z12627; NID:g14388; PIDN:CAA78274.1; PID:g14389
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992
C:Genetics:
A:Gene: atpA
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F;171-178/Region: nucleotide-binding motif A (P-loop)

F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>
F:268-273/Region: nucleotide-binding motif B
F:177/Binding site: ATP (lys) #status predicted
F:376/Active site: Arg #status predicted

Query Match 1.3%; Score 8; DB 1; Length 507;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
DB 236 AATASDPA 243

RESULT 38

H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - radish mitochondrion
C:Species: mitochondrion Raphanus sativus (radish)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C:Accession: S12309
R:Morikami, C.A.; Apel, I.J.; Palmer, J.D.
Plant Mol. Biol. 15, 735-746, 1990
A>Title: Characterization of radish mitochondrial atpA: influence of nuclear background
A:Reference number: S12309; MWID:91346710; PMID:2151720
A:Accession: S12309
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-507 <MAX>
A:Cross-references: UNIPROT:P23413
C:Genetics: atpA
A:Gene: atpA
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 507;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
DB 236 AATASDPA 243

RESULT 39

H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - garden pea mitochondrion
N:Alternate names: ATP synthase; ATPase; F1F0-ATPase alpha chain
C:Species: mitochondrion Pisum sativum (garden pea)
C>Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 09-Jul-2004
C:Accession: JN0769
R:Morikami, A.; Nakamura, K.
Biosci. Biotechnol. Biochem. 57, 1530-1535, 1993
A>Title: Transcript map of oppositely oriented pea mitochondrial genes encoding the alph
A:Reference number: JN0769; MWID:94033860; PMID:7764223
A:Accession: JN0769
A:Molecule type: DNA
A:Residues: 1-507 <MOR>
A:Cross-references: UNIPROT:P05493; GB:J14698; NID:g286139; PIDN:BA03524.1; PID:g286140
C:Comment: This enzyme is essential in the translocation of H+ through the inner mitoch
C:Genetics:
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 507;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
DB 236 AATASDPA 243

RESULT 40

H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - garden pea mitochondrion
C:Species: mitochondrion Pisum sativum (garden pea)
C>Date: 09-May-1988 #sequence_revision 09-May-1988 #text_change 09-Jul-2004
C:Accession: A26760
R:Morikami, A.; Nakamura, K.
J. Biochem. 101, 967-976, 1987
A>Title: Structure and expression of pea mitochondrial F1ATPase alpha-subunit gene and it
A:Reference number: A26760; MWID:87279992; PMID:2886497
A:Accession: A26760
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-507 <MOR>
A:Cross-references: UNIPROT:P05493
C:Genetics:
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 507;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
DB 236 AATASDPA 243

RESULT 41

H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - maize mitochondrion
N:Alternate names: ATPase alpha chain
C:Species: mitochondrion Zea mays (maize)
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: A23757
R:Braun, C.U.; Levings III, C.S.
Plant Physiol. 79, 571-577, 1985
A>Title: Nucleotide sequence of the F1-ATPase alpha subunit gene from maize mitochondria.
A:Reference number: A23757
A:Accession: A23757
A:Molecule type: DNA
A:Residues: 1-508 <BRA>
A:Cross-references: UNIPROT:P05494; EMBL:M16222; NID:g342633; PIDN:AAA70269.1; PID:g89761
C:Genetics:
A:Gene: atpA
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>
F:268-273/Region: nucleotide-binding motif B
F:177/Binding site: ATP (lys) #status predicted
F:376/Active site: Arg #status predicted

Query Match 1.3%; Score 8; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
DB 236 AATASDPA 243

RESULT 42

S29792

H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - soybean mitochondrion
N:Alternate names: ATPase alpha chain
C:Species: mitochondrion Glycine max (soybean)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C:Accession: S29792; S29873
R:Chanut, F.A.; Grabau, E.A.; Gesteland, R.F.
Curr. Genet. 23, 234-247, 1993
A:Title: Complex organization of the soybean mitochondrial genome: recombination repeats
A:Reference number: S29792; MUID:93169687; PMID:8435853
A:Accession: S29792
A:Molecule type: DNA
A:Residues: 1508 <CHA>
A:Cross-references: UNIPROT:Q01915; EMBL:Z14031; NID:G22738; PIDN:CAA79407.1; PID:G22739
A:Gene: atpA
C:Genetics:
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 508;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
DB 236 AATASDPA 243

RESULT 43
S26979
H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - kidney bean mitochondrion
N:Alternate names: ATPase alpha chain
C:Species: mitochondrion Phaseolus vulgaris (kidney bean)
C>Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 09-Jul-2004
C:Accession: S26979
R:Chase, C.D.; Ortega, V.M.
Curr. Genet. 22, 147-153, 1992
A:Title: Organization of ATPa coding and 3' flanking sequences associated with cytoplasm
A:Reference number: S26979; MUID:93046798; PMID:11423717
A:Accession: S26979
A:Molecule type: DNA
A:Residues: 1508 <CHA>
A:Cross-references: UNIPROT:P24459; EMBL:M64246; NID:9169317; PIDN:AA01582.1; PID:91693
A:Note: the authors translated the codon GCG for residue 257 as Cys
C:Genetics:
A:Gene: atpA
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 508;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
DB 236 AATASDPA 243

RESULT 44
PMTATC
H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - curled-leaved tobacco mito
C:Species: mitochondrion Nicotiana glauca (curled-leaved tobacco)
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: S00956
R:Chamont, F.; Boutry, M.; Briquet, M.; Vassarotti, A.
Nucleic Acids Res. 16, 6247, 1988
A:Title: Sequence of the gene encoding the mitochondrial F1-ATPase alpha subunit from Ni
A:Reference number: S00956; MUID:88289387; PMID:2899872

A:Accession: S00956
A:Molecule type: DNA
A:Residues: 1509 <CHA>
A:Cross-references: UNIPROT:P05495; EMBL:X07745; NID:G13152; PIDN:CAA30568.1; PID:G13153
C:Genetics:
A:Gene: atpA
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>
F:268-273/Region: nucleotide-binding motif B
F:177/Binding site: ATP (Lys) #status predicted
F:376/Active site: Arg #status predicted

Query Match 1.3%; Score 8; DB 1; Length 509;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
DB 236 AATASDPA 243

RESULT 45
PWRZAM
H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - rice mitochondrion
C:Species: mitochondrion Oryza sativa (rice)
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: JQ0411
R:Kadowaki, K.; Kazama, S.; Suzuki, T.
Nucleic Acids Res. 18, 1302, 1990
A:Title: Nucleotide sequence of the F1-ATPase alpha subunit gene from rice mitochondria.
A:Reference number: JQ0411; MUID:90206808; PMID:2138730
A:Accession: JQ0411
A:Molecule type: DNA
A:Residues: 1509 <KAD>
A:Cross-references: UNIPROT:P15998; EMBL:X51422; NID:G13958; PIDN:CAA35787.1; PID:G13959
C:Genetics:
A:Gene: atpA
A:Genome: mitochondrion
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>
F:268-273/Region: nucleotide-binding motif B
F:177/Binding site: ATP (Lys) #status predicted
F:376/Active site: Arg #status predicted

Query Match 1.3%; Score 8; DB 1; Length 509;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
DB 236 AATASDPA 243

RESULT 46
PMTATM
H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - wheat mitochondrion
C:Species: mitochondrion Triticum aestivum (common wheat)
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: S06007
R:Schulte, E.; Stauch, S.; Iaser, B.; Kueck, U.
Nucleic Acids Res. 17, 7531, 1989
A:Title: Wheat mitochondrial DNA: organization and sequences of the atpA and atp9 genes.
A:Reference number: S06007; MUID:90016824; PMID:2529479
A:Accession: S06007
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1509 <SCH>
A:Cross-references: UNIPROT:P12862; EMBL:X15918

C/Genetics:
A:Gene: atpA
A:Genome: mitochondrion
C/Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C/Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>
F:268-273/Region: nucleotide-binding motif B
F:177/Binding site: ATP (Lys) #status predicted
F:376/Active site: Arg #status predicted

Query Match 1.3%; Score 8; DB 1; Length 509;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
Db 236 AATASDPA 243

RESULT 47
T11937
H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - Prototheca wickerhamii mit
N:Alternate names: H(+)-transporting ATPase, subunit 1
C/Species: mitochondrion Prototheca wickerhamii
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C/Accession: T11937
R/Wolff, G.; Planke, I.; Lang, B.F.; Kueck, U.; Burger, G.
J. Mol. Biol. 237, 75-86, 1994
A>Title: Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca w
A/Reference number: Z17373; MUID:94180393; PMID:8133522
A/Accession: T11937
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-509 <WOL>
A/Cross-references: UNIPROT:Q37628; EMBL:U02970; NID:9467843; PID:9467869; PIDN:AAD12656
A/Experimental source: strain 263-11
C/Genetics:
A:Gene: atp1
A:Genome: mitochondrion
C/Function:
A/Pathway: oxidative phosphorylation
C/Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C/Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxi
F:202-374/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 509;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
Db 233 AATASDPA 240

RESULT 48
H97673
ATP synthase alpha chain [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C/Accession: H97673
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: H97673
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-509 <KUR>
A/Cross-references: UNIPROT:Q8UC74; GB:AE007869; PIDN:AAK8345.1; PID:915157826; GSPDB:C
C/Genetics:
A:Gene: AGR_C_4757

A/Map position: circular chromosome
C/Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al

Query Match 1.3%; Score 8; DB 2; Length 509;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
Db 233 AATASDPA 240

RESULT 49
AG2898
ATP synthase alpha chain atpA [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C/Accession: AG2898
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AG2898
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-509 <KUR>
A/Cross-references: UNIPROT:Q8UC74; GB:AE008688; PIDN:AA43605.1; PID:917741124; GSPDB:GT
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A:Gene: atpA
A/Map position: circular chromosome
C/Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al

Query Match 1.3%; Score 8; DB 2; Length 509;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
Db 233 AATASDPA 240

RESULT 50
AD3283
H+-transporting two-sector ATPase (EC 3.6.3.14) [imported] - Brucella melitensis (strain
C/Species: Brucella melitensis
C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C/Accession: AD3283
R/DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I
; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesc
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A/Reference number: AD3252; PMID:11756688
A/Accession: AD3283
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-509 <KUR>
A/Cross-references: UNIPROT:Q8YJ37; GB:AE008917; PIDN:AA151431.1; PID:917982138; GSPDB:GT
A/Experimental source: strain 16M
C/Genetics:
A:Gene: BME10249
A/Map position: 1
C/Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al

Query Match 1.3%; Score 8; DB 2; Length 509;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
 |||||
 DB 233 AATASDPA 240

RESULT 51

S19261
 H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - common sunflower mitochondrion
 C/Species: mitochondrion Helianthus annuus (common sunflower)
 C/Date: 19-Feb-1994 #sequence_revision 03-Aug-1995 #text_change 09-Jul-2004
 A/Accession: S19261
 R/Stcullella, L.; D'Ambrosio, L.; de Tuglie, A.D.; Gallerani, R.
 Nucleic Acids Res. 18, 4599, 1990
 A/Title: Minor differences in the primary structures of atpA genes coded on the mtDNA of
 A/Reference number: S10974; MUID:90356407; PMID:2143818
 A/Accession: S19261
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-510 <SIC>
 A/Cross-references: UNIPROT:P18260; EMBL:X52838
 A/Experimental source: strain CMS89
 C/Genetics:
 A/Genome: atpA
 A/Genome: mitochondrion
 C/Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
 C/Keywords: ATP biosynthesis; hydrolase; mitochondrion; nucleotide binding; P-loop
 F/171-178/Region: nucleotide-binding motif A (P-loop)
 F/205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 510;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
 |||||
 DB 236 AATASDPA 243

RESULT 52

S10997
 H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - common sunflower mitochondrion
 C/Species: mitochondrion Helianthus annuus (common sunflower)
 C/Date: 18-Feb-1994 #sequence_revision 03-Aug-1995 #text_change 09-Jul-2004
 A/Accession: S10997; S10974; S17799; S52010; S48854
 R/Kochler, R.H.; Loessel, A.; Zetsche, K.
 Nucleic Acids Res. 18, 4588, 1990
 A/Title: Nucleotide sequence of the F1-ATPase alpha subunit gene of sunflower mitochondrion
 A/Reference number: S10997; MUID:90356396; PMID:2143817
 A/Accession: S10997
 A/Molecule type: DNA
 A/Residues: 1-510 <XOE>
 A/Cross-references: UNIPROT:P18260; EMBL:X53537; NID:g12577; PIDN:CAA37613.1; PID:g75836
 A/Experimental source: line Baso
 R/Stcullella, L.; D'Ambrosio, L.; de Tuglie, A.D.; Gallerani, R.
 Nucleic Acids Res. 18, 4599, 1990
 A/Title: Minor differences in the primary structures of atpA genes coded on the mtDNA of
 A/Reference number: S10974; MUID:90356407; PMID:2143818
 A/Accession: S10974

A/Molecule type: DNA
 A/Residues: 1-510 <SIC>
 A/Cross-references: EMBL:X52838; NID:g12988; PIDN:CAA37022.1; PID:g12969
 A/Experimental source: strain HA89
 R/Kochler, R.H.; Horn, R.; Loessel, A.; Zetsche, K.
 Mol. Gen. Genet. 227, 369-376, 1991
 A/Title: Cytoplasmic male sterility in sunflower is correlated with the co-transcription
 A/Reference number: S16528; MUID:91326025; PMID:1714033
 A/Accession: S17799
 A/Status: translation not shown
 A/Molecule type: DNA
 A/Residues: 1-510 <KOA>
 A/Cross-references: EMBL:X53963; NID:g12985; PIDN:CAA39428.1; PID:g12986
 A/Experimental source: line CMSBaso
 R/Spassova, M.; Moninger, F.; Leaver, C.J.; Petrov, P.; Atanassov, A.; Nijkamp, H.J.J.; H

Plant Mol. Biol. 26, 1819-1831, 1994
 A/Title: Characterisation and expression of the mitochondrial genome of a new type of cy
 A/Reference number: S52010; MUID:95161706; PMID:7858220
 A/Accession: S52010
 A/Status: translation not shown

A/Molecule type: DNA
 A/Residues: 293-510 <SPA>
 A/Cross-references: EMBL:X82386; NID:g563544; PIDN:CAA57786.1; PID:g563545
 A/Genetics:
 A/Genome: atpA
 A/Genome: mitochondrion
 C/Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
 C/Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc
 F/171-178/Region: nucleotide-binding motif A (P-loop)
 F/205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 510;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
 |||||
 DB 236 AATASDPA 243

RESULT 53

S07316
 H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - German evening primrose m
 C/Species: mitochondrion Oenothera biennis (German evening primrose)
 C/Date: 21-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 09-Jul-2004
 A/Accession: S07316
 R/Schuster, W.; Brennicke, A.
 Mol. Gen. Genet. 204, 29-35, 1986
 A/Title: Pseudocopies of the ATPase alpha-subunit gene in Oenothera mitochondria are pre
 A/Reference number: S07316
 A/Accession: S07316
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-511 <SCH>
 A/Cross-references: UNIPROT:P05492; EMBL:X04023; NID:g13164; PIDN:CAA27656.1; PID:g13165
 C/Genetics:
 A/Genome: mitochondrion
 C/Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
 C/Keywords: ATP biosynthesis; hydrolase; mitochondrion; nucleotide binding; P-loop
 F/171-178/Region: nucleotide-binding motif A (P-loop)
 F/205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 511;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
 |||||
 DB 236 AATASDPA 243

RESULT 54

S04672
 H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - Rhodospseudomonas blastic
 C/Species: Rhodospseudomonas blastica
 C/Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
 A/Accession: S04672
 R/Tybuliewicz, V.L.J.; Falk, G.; Walker, J.E.
 J. Mol. Biol. 179, 185-214, 1984
 A/Title: Rhodospseudomonas blastica atp operon. Nucleotide sequence and transcription.
 A/Reference number: S04666; MUID:85058188; PMID:6209404
 A/Accession: S04672
 A/Status: not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 1-512 <TYB>
 A/Cross-references: UNIPROT:P05439
 C/Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
 C/Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; oxidative phosphory

F:204-376/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 1.3%; Score 8; DB 2; Length 512;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
|||||
DB 235 AATASDPA 242

RESULT 55

S25955

H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - liverwort (Marchantia poly-

C:Species: Marchantia polymorpha

C>Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 09-Jul-2004

C:Accession: S25955

R:Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takemura, M.; Nozato, N.; Akashi, K.; Kat

J. Mol. Biol. 223, 1-7, 1992

A>Title: Gene organization deduced from the complete sequence of liverwort Marchantia po

A:Reference number: S25941; PMID:1731062

A:Accession: S25955

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-513 <ODA>

A:Cross-references: UNIPROT:P26854; EMBL:M68929; NID:9786182; PIDN:AC09446.1; PID:97862

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992

C:Genetics:

A:Gene: atpA

A:Genome: mitochondrion

A:Introns: 330/2; 350/3

C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al

C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; nuc

F:170-177/Region: nucleotide-binding motif A (P-loop)

F:204-376/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

QY 163 AATASDPA 170
|||||
DB 235 AATASDPA 242

RESULT 56

T14604

hypothetical 59.6K protein - Trypanosoma cruzi

C:Species: Trypanosoma cruzi

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C:Accession: T14604

R:Andersson, B.; Aslund, L.; Pettersson, U.

submitted to the EMBL Data Library, March 1998

A:Description: 93.4 kb of complete sequence from chromosome 3 of Trypanosoma cruzi.

A:Reference number: Z18159

A:Accession: T14604

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-546 <AND>

A:Cross-references: UNIPROT:O61097; EMBL:AF052832; NID:93063540; PID:93063542; PIDN:AC0

C:Genetics:

A:Map position: 3

C:Superfamily: Trypanosoma cruzi hypothetical 59.6K protein

QY 27 RAALGCP 34
|||||
DB 220 RAALGCP 227

Query Match 1.3%; Score 8; DB 2; Length 546;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 57

T22341

hypothetical protein F47B8.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T22341

R:Berke, M.; McMurtry, A.

submitted to the EMBL Data Library, July 1996

A:Reference number: Z19551

A:Accession: T22341

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-581 <WIL>

A:Cross-references: UNIPROT:Q20517; EMBL:Z77662; PIDN:CB01192.1; GSPDB:GN00023; CESP:F4

A:Experimental source: clone F47B8

C:Genetics:

A:Gene: CESP:F47B8.5

A:Map position: 5

A:Introns: 37/3; 82/3; 122/3; 146/3; 194/1; 217/3; 298/2; 368/1

Query Match 1.3%; Score 8; DB 2; Length 581;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAGG 17
|||||
DB 451 AAAAPAGG 458

RESULT 58

A26638

homeotic protein Dfd - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jul-2004

C:Accession: A26638; B26638; A24780

R:Regulski, M.; McGinnis, N.; Chadwick, R.; McGinnis, W.

EMBO J. 6, 767-777, 1987

A>Title: Developmental and molecular analysis of Deformed: a homeotic gene controlling D

A:Reference number: A26638

A:Accession: A26638

A:Molecule type: DNA

A:Residues: 1-590 <REL>

A:Cross-references: UNIPROT:P07548

A:Experimental source: strain Canton S

A>Note: the authors translated the codon GTC for residue 32 as Leu

A:Accession: B26638

A:Molecule type: mRNA

A:Residues: 1-207; 'K', 209-477; 481-544; 546-590 <RE2>

A:Experimental source: strain Oregon R

R:Regulski, M.; Harding, K.; Kostriken, R.; Karch, F.; Levine, M.; McGinnis, W.

Cell 43, 71-80, 1985

A>Title: Homeo box genes of the Antennapedia and Bithorax complexes of Drosophila.

A:Reference number: A90874; PMID:86079516; PMID:2416463

A:Accession: A24780

A:Molecule type: DNA

A:Residues: 366-426 <RE3>

C:Genetics:

A:Gene: FlyBase:Dfd

A:Cross-references: FlyBase:FBgn000439

C:Superfamily: unassigned homeobox proteins; homeobox homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:367-423/Domain: homeobox homology <HOX>

Query Match 1.3%; Score 8; DB 2; Length 590;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGA 215
|||||
DB 189 GGGGSGA 196

RESULT 59

A53731
 translation initiation factor eIF-2 alpha chain kinase (EC 2.7.1.-) - rat
 N/Alternate names: heme-controlled repressor
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 09-Jul-2004
 C/Accession: A53731
 R/Mellor, H.; Flowers, K.M.; Kimball, S.R.; Jefferson, L.S.
 J. Biol. Chem. 269, 10201-10204, 1994
 A/Title: Cloning and characterization of cDNA encoding rat hemin-sensitive initiation factor
 A/Reference number: A53731; MUID:94193700; PMID:7908250
 A/Accession: A53731
 A/Status: preliminary; not compared with conceptual translation
 A/Molecule type: mRNA
 A/Residues: 1-620 <MEL>
 A/Cross-references: UNIPROT:Q63185; GB:L27707; NID:9443688; PID:AAA18255.1; PID:9443688
 C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
 C/Keywords: phosphotransferase
 F:165-581/Domain: protein kinase homology <KIN>

Query Match 1.3%; Score 8; DB 2; Length 620;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 EKEIEELK 251
 |||||
 Db 599 EKEIEELK 606

RESULT 60
 A36353
 DNA repair protein XRCC1 - human
 C/Species: Homo sapiens (man)
 C/Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 09-Jul-2004
 C/Accession: A36353
 R/Thompson, L.H.; Brookman, K.W.; Jones, N.J.; Allen, S.A.; Carrano, A.V.
 Mol. Cell. Biol. 10, 6160-6171, 1990
 A/Title: Molecular cloning of the human XRCC1 gene, which corrects defective DNA strand
 A/Reference number: A36353; MUID:91061722; PMID:2247054
 A/Accession: A36353
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-633 <TRD>
 A/Cross-references: UNIPROT:P18887; GB:M36089; NID:9340396; PID:AAA63270.1; PID:9340397
 C/Genetics:
 A/Genes: GDB:XRCC1; RCC
 A/Cross-references: GDB:120737; OMIM:194360
 A/Map position: 19q13.2-19q13.2

Query Match 1.3%; Score 8; DB 2; Length 633;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 TADDPAGP 172
 |||||
 Db 202 TADDPAGP 209

RESULT 61
 A55719
 dnaK-type molecular chaperone HSPA2 - human
 N/Alternate names: heat shock protein A2
 C/Species: Homo sapiens (man)
 C/Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 09-Jul-2004
 C/Accession: A55719
 R/Bonnycaz, L.L.C.; Yu, C.E.; Hunt, C.R.; Trask, B.J.; Clancy, K.P.; Weber, J.L.; Pat
 Genomics 23, 85-93, 1994
 A/Title: Cloning, sequencing, and mapping of the human chromosome 14 heat shock protein
 A/Reference number: A55719; MUID:9510116; PMID:7829106
 A/Accession: A55719
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-639 <BON>
 A/Cross-references: UNIPROT:P54652; GB:L26336; NID:9476704; PID:AAA52698.1; PID:9476705

C/Genetics:
 A/Genes: GDB:HSPA2
 A/Cross-references: GDB:120059; OMIM:140560
 A/Map position: 14q22-14q22
 A/Introns: #status absent
 C/Function:
 A/Description: involved in protein folding and assembling/disassembling of protein complex
 C/Superfamily: heat shock protein 70
 C/Keywords: ATP; molecular chaperone

Query Match 1.3%; Score 8; DB 2; Length 639;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 GGGGSGAS 216
 |||||
 Db 623 GGGGSGAS 630

RESULT 62
 A43800
 nuclear autoantigenic sperm protein - rabbit
 C/Species: Oryctolagus cuniculus (domestic rabbit)
 C/Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
 C/Accession: A43800
 R/Melch, J.E.; Zimmerman, L.J.; Joseph, D.R.; O'Rand, M.G.
 Biol. Reprod. 43, 559-568, 1990
 A/Title: Characterization of a sperm-specific nuclear autoantigenic protein. Complete seq
 A/Reference number: A43800; MUID:91145522; PMID:2289010
 A/Accession: A43800
 A/Molecule type: mRNA; DNA
 A/Residues: 1-680 <MEL>
 A/Cross-references: UNIPROT:P27123; GB:M37893; NID:9165554; PID:AAA31423.1; PID:9165555

Query Match 1.3%; Score 8; DB 2; Length 680;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 EKEIEELK 251
 |||||
 Db 529 EKEIEELK 536

RESULT 63
 T01209
 starch synthase (EC 2.4.1.21) isoform STSII-2 - maize
 N/Alternate names: starch synthase isoform STSII-2
 C/Species: Zea mays (maize)
 C/Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 09-Jul-2004
 C/Accession: T01209
 R/Knight, M.E.; Harn, C.; Lilley, C.E.R.; Guan, H.P.; Singletary, G.W.; Mu-Forester, C.;
 Plant J. 14, 613-622, 1998
 A/Title: Molecular cloning of starch synthase I from maize (W64A) endosperm and expressi
 A/Reference number: T14279; MUID:98340555; PMID:9675904
 A/Accession: T01209
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-698 <KNT>
 A/Cross-references: UNIPROT:O48900; EMBL:AF019297; NID:92655030; PID:AA013342.1; PID:926
 A/Experimental source: strain W64A; endosperm
 C/Genetics:
 A/Genes: SSIIb
 C/Function:
 A/Description: catalyzes the alpha-1,4-glucosylation of starch by ADPGlucose producing e)
 C/Superfamily: starch synthase
 C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 1.3%; Score 8; DB 2; Length 698;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAGC 17
 |||||

Db 61 AAAAPAGG 68

RESULT 64

S28030
DNA-binding protein Gt-2 - rice
C:Species: Oryza sativa (rice)
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C:Accession: S28030; A57288
R:Dehesh, K.; Hung, H.; Tepperman, J.M.; Quail, P.H.
EMBO J. 11, 4131-4144, 1992
A>Title: Gt-2: a transcription factor with twin autonomous DNA-binding domains of closed
A:Reference number: S28030; NUID:93011008; PMID:1196594
A:Accession: S28030
A:Molecule type: DNA
A:Residues: 1-737 <DEH>
A:Cross-references: UNIPROT:Q02227; EMBL:X68261; NID:920248; PIDN:CAA8328.1; PID:920249
R:Dehesh, K.; Bruce, W.B.; Quail, P.H.
Science 250, 1397-1399, 1990
A>Title: A trans-acting factor that binds to a GT-motif in a phytochrome gene promoter.
A:Reference number: A37288; NUID:91075225; PMID:2255908
A:Accession: A37288
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: mRNA
A:Residues: 404-737 <DE2>

Query Match 1.3%; Score 8; DB 2; Length 737;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGA 215
Db 87 GGGGSGGA 94

RESULT 65

S61247
DNA helicase/primase complex associated protein - bovine herpesvirus 1
C:Species: bovine herpesvirus 1
C>Date: 18-Sep-1997 #sequence_revision 18-Sep-1997 #text_change 09-Jul-2004
C:Accession: S61247
R:Vlack, C.; Benes, V.; Lu, Z.; Kutish, G.F.; Paces, V.; Rock, D.; Lechworth, G.J.; Sch
submitted to the EMBL Data Library, January 1995
A:Description: Nucleotide sequence analysis of a 30-kb region of the bovine herpesvirus
A:Reference number: S61233
A:Accession: S61247
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-748 <VLC>
A:Cross-references: UNIPROT:P52374; EMBL:Z48053; NID:9971311; PIDN:CAA88125.1; PID:99713
C:Superfamily: varicella-zoster virus gene 52 protein

Query Match 1.3%; Score 8; DB 2; Length 748;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAGG 17
Db 165 AAAAPAGG 172

RESULT 66

S54016
SOK2 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YMR911.03c; protein YMR016c
C:Species: Saccharomyces cerevisiae
C>Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: S54016
R:Lyte, G.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54014
A:Accession: S54016
A:Molecule type: DNA

A:Residues: 1-785 <LYE>
A:Cross-references: UNIPROT:P53438; EMBL:Z49211; NID:9798922; PID:9798925; MIPS:YMR016c
A:Experimental source: strain AB972
C:Genetics:
A:Gene: SGD:SOK2
A:Cross-references: SGD:S0004618; MIPS:YMR016c
A:Map position: 13R

Query Match 1.3%; Score 8; DB 2; Length 785;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 OOOOOLQA 242
Db 211 OOOOOLQA 218

RESULT 67

B41054
fasciclin II PI-linked splice form precursor - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 17-Mar-2000
C:Accession: B41054
R:Gemingloh, G.; Rehm, B.J.; Goodman, C.S.
Cell 67, 45-57, 1991
A>Title: Genetic analysis of growth cone guidance in Drosophila: fasciclin II functions e
A:Reference number: A41054; NUID:92005695; PMID:1913618
A:Accession: B41054
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-873 <GRE>
A:Cross-references: GB:M77166
C:Genetics:
A:Gene: FlyBase:Fab2
A:Cross-references: FlyBase:FBgn0000635
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C:Keywords: transmembrane protein

Query Match 1.3%; Score 8; DB 2; Length 873;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 PPSPLPL 446
Db 811 PPSPLPL 818

RESULT 68

T49801
hypothetical protein B11B22.30 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C:Accession: T49801
R:Schulte, U.; Algm, V.; Hobeisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49801
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-876 <SCH>
A:Cross-references: UNIPROT:Q9P4Y9; EMBL:AL356834; GSPDB:GN00116; NCSP:B11B22.30
A:Experimental source: BAC clone B11B22; strain OR74A
C:Genetics:
A:Gene: NCSP:B11B22.30
A:Map position: 6
A:Intons: 75/3; 190/1; 449/3

Query Match 1.3%; Score 8; DB 2; Length 876;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 QPPSPPLP 445

Db 357 QPPSPRP 364

RESULT 69

T01269

serine/threonine-specific protein kinase (EC 2.7.1.-) F27F23.1 - Arabidopsis thaliana

N/Alternate names: protein T20K24.22; receptor-like protein kinase

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Aug-2004

C/Accession: T01269; T00542; H84573

R/Rounley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul

submitted to the EMBL Data Library, May 1998

A/Description: Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence.

A/Reference number: Z14177

A/Accession: T01269

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-881 <ROU>

A/Cross-references: UNIPROT:O65924; EMBL:AC003058; NID:g3135250; PIDN:AAC16451.1; PID:g3

A/Experimental source: cultivar Columbia

R/Rounley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,

submitted to the EMBL Data Library, July 1997

A/Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.

A/Reference number: Z14167

A/Accession: T00542

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-881 <RO2>

A/Cross-references: EMBL:AC002392; NID:g3176701; PIDN:AAD12034.1; PID:g3176719

A/Experimental source: cultivar Columbia

R/Rin, X.; Kaul, S.; Rounley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Unayam, L.; Tallon, L.

gens, D.; Niemann, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J

Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: H84573

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-881 <STO>

A/Cross-references: GB:AE002093; NID:g3135251; PIDN:AAC16451.1; GSPDB:GN00139

C/Genetics:

A/Gene: F27F23.1; T20K24.22; AT2G19210

A/Map position: 2

A/Intons: 30/1; 212/2; 372/1; 418/2; 442/2; 466/2; 493/2; 549/1; 619/3; 662/1; 683/1; 7

C/Superfamily: protein kinase homology

C/KeyWord: phosphotransferase

C/Superfamily: bromodomain homology

F/48-136/Domain: RING finger homology <REN>

F/925-981/Domain: bromodomain homology <BRO1>

Query Match 1.3%; Score 8; DB 2; Length 1051;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 10 AAAAPAGC 17

10 AAAAPAGC 17

RESULT 71

A56155

tumor suppressor protein warts (EC 2.7.1.-) - fruit fly (Drosophila melanogaster)

C/Species: Drosophila melanogaster

C/Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 09-Jul-2004

C/Accession: A56155

R/Justice, R.W.; Zilian, O.; Woode, D.F.; Noll, M.; Bryant, P.J.

Genes Dev. 9, 534-546, 1995

A/Title: The Drosophila tumor suppressor gene warts encodes a homolog of human myotonic

A/Reference number: A56155; MUID:55212904; PMID:7658644

A/Accession: A56155

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1099 <JUS>

A/Cross-references: UNIPROT:Q24590; GB:J39837; NID:g755007; PIDN:AA73959.1; PID:g755008

C/Genetics:

A/Gene: wts

A/Cross-references: FlyBase:FBgn0011739

C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C/KeyWord: ATP; phosphotransferase

F/11-1014/Domain: protein kinase homology <KIN>

F/719-727/Region: protein kinase ATP-binding motif

Query Match 1.3%; Score 8; DB 2; Length 1099;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 235 QQQQQLQA 242

498 QQQQQLQA 505

RESULT 72

S25954

gene atpa intron 2 protein - liverwort (Marchantia polymorpha) mitochondrion

C/Species: Marchantia polymorpha

C/Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 09-Jul-2004

C/Accession: S25954

R/Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takemura, M.; Nozato, N.; Akashi, K.; Kan

J. Mol. Biol. 223, 1-7, 1992

A/Title: Gene organization deduced from the complete sequence of liverwort Marchantia po

A/Reference number: S25941; MUID:92114051; PMID:1731062

A/Accession: S25954

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-1259 <ODA>

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992

C/Genetics:

A/Genome: mitochondrion

A/Intons: 330/2

C/Superfamily: gene atpa intron protein; H⁺-transporting ATP synthase alpha chain homolo

C/KeyWord: mitochondrion; nucleotide binding; P-loop

F/1-350/Region: atpa exons 1 and 2 encoded

F/110-177/Region: nucleotide-binding motif A (P-loop)

F/351-1259/Region: atpa intron encoded

Query Match 1.3%; Score 8; DB 2; Length 1259;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
|||||
DB 235 AATASDPA 242

RESULT 73

B66292
F/H2.12 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: B66292
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Com, D.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, D.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
A:Authors: Selzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; WUID:21016719; PMID:11130712
A:Accession: B66292
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1366 <STO>
A:Cross-references: UNIPROT:Q9JLM06; GB:AE005172; NID:98927657; PIDN:AAF82148.1; GSPDB:GN
C:Genetics:
A:Map position: 1

Query Match 1.3%; Score 8; DB 2; Length 1366;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 000000LOA 242
|||||
DB 891 000000LOA 898

RESULT 74

S25997
gene atpa intron 1 protein - liverwort (Marchantia polymorpha) mitochondrion
C:Species: mitochondrion Marchantia polymorpha
C>Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 09-Jul-2004
C:Accession: S25997
R:Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takemura, M.; Nozato, N.; Akashi, K.; Kan
J. Mol. Biol. 223, 1-7, 1992
A>Title: Gene organization deduced from the complete sequence of liverwort Marchantia po
A:Reference number: S25941; WUID:92114051; PMID:1731062
A:Accession: S25997
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1395 <ODA>
A:Cross-references: UNIPROT:Q35059; EMBL:M66929; NID:9786182; PIDN:AAC09444.1; PID:97862
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992
C:Genetics:
A:Genome: mitochondrion
C:Superfamily: gene atpa intron protein; H+-transporting ATP synthase alpha chain homolo
C:Keyword: mitochondrion; nucleotide binding; P-loop
F:1-329/Region: atpa exon 1 encoded
F:170-177/Region: nucleotide-binding motif A (P-loop)
F:331-1395/Region: atpa intron encoded

Query Match 1.3%; Score 8; DB 2; Length 1395;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
|||||
DB 235 AATASDPA 242

RESULT 75

T08428
gene small optic lobes protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T08428
R:Maliszka, R.; de Couet, H.G.; Miklos, G.L.
Proc. Natl. Acad. Sci. U.S.A. 95, 3731-3736, 1998
A>Title: Data transferability from model organisms to human beings: insights from the fur
A:Reference number: Z16415; WUID:98188272; PMID:9520435
A:Accession: T08428
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1597 <MAL>
A:Cross-references: UNIPROT:O61346; EMBL:AF017777; NID:93004652; PIDN:AAC28409.1; PID:930
A:Experimental source: strain Cantons
C:Genetics:
A:Gene: sol
A:Cross-references: FlyBase:FBgn0024251
A:Insertions: 81/2; 290/1; 594/2; 704/1; 1233/3; 1402/3; 1491/3; 1539/2
C:Superfamily: sol protein; calpain catalytic domain homology
F:1047-1307/Domain: calpain catalytic domain homology <CALP>

Query Match 1.3%; Score 8; DB 2; Length 1597;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 EQ0000LOQ 241
|||||
DB 672 EQ0000LOQ 679

Search completed: March 23, 2005, 15:52:44
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 25, 2005, 21:09:28 ; Search time 288 Seconds
(without alignments)
3488.451 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 3238
Sequence: 1 MTRSAVFKAAPAGAGNPE.....RSRCRLIEIQKQTPHRYCRK 614

Scoring table:
BLASTSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=x1h
-Q=/cgn2_1/USFTO.spool/US10054935/runat_23032005_144820_26113/app_query.fasta_1.775
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=rml -MINMATCH=0.1 -LOOFCU=0
-LOOPEXT=0 -UNITS=b16s -START=1 -END=-1 -MATRIX=b16sum62 -TRANS=human40.cdi
-LIST=45 -DOCLINK=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=plo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US10054935 @CEN 1.1 69 @runat_23032005_144820_26113 -NCPU=6 -ICPU=3
-NO_WAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WAPN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1053.5	32.5	1290	4	US-09-023-655-322 Sequence 322, Appl
2	238	7.4	9551	4	US-08-056-200-93 Sequence 93, Appl
3	238	7.4	9551	2	US-08-800-644-93 Sequence 93, Appl
4	230.5	7.1	7065	4	US-09-874-923-115 Sequence 115, Appl
5	230	7.1	20076	4	US-09-949-016-13141 Sequence 13141, A
6	230	7.1	25973	4	US-09-949-016-14140 Sequence 14140, A
7	228.5	7.1	2730	4	US-09-902-540-489 Sequence 489, Appl
8	228.5	7.1	3169	4	US-09-620-312D-713 Sequence 713, Appl
9	224.5	6.9	14555	4	US-09-902-540-1096 Sequence 1096, Appl
10	224	6.9	5788	4	US-09-949-016-12458 Sequence 12458, A
11	224	6.9	5788	4	US-09-949-016-14458 Sequence 14458, A
12	224	6.9	5870	4	US-09-949-016-15247 Sequence 15247, A

13	223.5	6.9	3117	4	US-09-081-385-6 Sequence 6, Appli
14	223	6.9	16562	4	US-09-949-016-13892 Sequence 12892, A
15	222.5	6.9	24905	4	US-09-902-540-1225 Sequence 1225, Ap
16	222	6.9	3168	4	US-09-902-540-8270 Sequence 8270, Ap
17	222	6.9	8050	4	US-09-902-540-855 Sequence 855, App
18	221.5	6.8	5883	4	US-09-949-016-5001 Sequence 5001, Ap
19	220	6.8	2841	4	US-09-902-540-9681 Sequence 9681, Ap
20	220	6.8	13637	4	US-09-902-540-1097 Sequence 1097, Ap
21	220	6.8	21105	4	US-09-949-016-15722 Sequence 15722, A
22	220	6.8	154746	4	US-09-827-688-8 Sequence 8, Appli
23	220	6.8	154746	4	US-09-827-688-8 Sequence 8, Appli
24	219.5	6.8	8048	4	US-09-902-540-867 Sequence 867, App
25	217	6.7	36778	3	US-09-105-537-5 Sequence 5, Appli
26	217	6.7	38506	3	US-09-320-878-19 Sequence 19, Appli
27	217	6.7	38506	4	US-09-141-908-1 Sequence 1, Appli
28	217	6.7	38506	4	US-09-657-440-19 Sequence 19, Appli
29	216	6.7	2862	4	US-09-902-540-7479 Sequence 7479, Ap
30	216	6.7	5482	4	US-09-902-540-723 Sequence 723, App
31	215.5	6.7	3147	2	US-08-781-802-7 Sequence 7, Appli
32	215.5	6.7	3147	3	US-08-694-078-7 Sequence 7, Appli
33	215.5	6.7	3147	3	US-09-058-260-7 Sequence 7, Appli
34	215.5	6.7	3568	4	US-09-902-540-4681 Sequence 4681, Ap
35	215.5	6.7	6644	4	US-08-875-435B-5 Sequence 5, Appli
36	215.5	6.7	24754	4	US-09-902-540-1230 Sequence 1230, Ap
37	215	6.6	1632	4	US-09-902-540-5283 Sequence 5283, Ap
38	215	6.6	9080	4	US-09-902-540-1038 Sequence 1038, Ap
39	215	6.6	11706	4	US-09-902-540-1038 Sequence 1110, Ap
40	215	6.6	15209	4	US-09-902-540-3216 Sequence 3216, App
41	214.5	6.6	1938	4	US-09-902-540-952 Sequence 952, App
42	214.5	6.6	9992	4	US-09-902-540-1147 Sequence 1147, Ap
43	214.5	6.6	19455	4	US-09-902-540-1147 Sequence 1, Appli
44	214.5	6.6	4411529	3	US-09-103-840A-1 Sequence 11951, A
45	213.5	6.6	7890	4	US-09-949-016-11951

ALIGNMENTS

RESULT 1
US-09-023-655-322
Sequence 322, Application US/09023655
Patent No. 6607879
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
APPLICANT: Susan G. Stuart
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESSES:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,655
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 322:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMNOR01
; CLONE: 140704
; US-09-023-655-322

Alignment Scores:
Pred. No.: 6,84e-49 Length: 1290
Score: 1053.50 Matches: 207
Percent Similarity: 93.69% Conservative: 1
Best Local Similarity: 93.24% Mismatches: 6
Query Match: 32.54% Indels: 8
DB: 4 Gaps: 2

US-10-054-935-2 (1-614) x US-09-023-655-322 (1-1290)
QY 258 LeuAlArGlIleGluArGMeTGlUArGArGMeTGlUleuValIySlySAsPaNgIuLys 277
DB 3 CTTGCTCGATTGAAGCTATGAAAGCGGATGCACTGTTAAAGAAAGATTAACGAGAAA 62
QY 278 GlUArGHisIySleuPheGInGlyTyrgIuThrGluGluArGluGluThrGluLeuSer 297
DB 63 GAAAGGCACAAAGCTGTTTCAGGGCTATGAAACTGAAAGAGAGAGAGAGAAACAGAGCTATCT 122
QY 298 GlUlySleIleIySleuGluCySgInPProGluLeuSerGluThrSerGluThrLeuProPro 317
DB 123 GAAATAATTAACTGAGTGCCAGCCGAGCTTTCAGAGCATCCCGAGCTGCTGCCCTCC 182
QY 318 LysPProPheSerCySglYArGser-GlyUySgIyHisIyArGlyuSerProPheGlySe 337
DB 183 AAGCCCTTCTCATGTGGCGGAGGGGAAAGGACATAAAGAAATCCCATTTGGAG 242
QY 337 rThrGluArGlySleThrProValIySlySleuAlaProGluPheSerIySValIySThrIy 357
DB 243 TACAGAAAGAAAGACTCTCTGTTAAAAAGCTGGCTCCGATTTTCAAAAGTCAAAACAAA 302
QY 357 sThrPProIyHisIySerProIleIySgIuGluProCySgIySerIySleuSerGluThrVal 377
DB 303 AACTCTTAAGCACTCTCTATTAAAGAGAAACCTTGCTTCTTAATCTGAAACTGTTTG 362
QY 377 sIySArGglUleuArGserGInGluThrPProGluIySProArGserSerAlaSerThrPr 397
DB 363 TAAACGTGAATTAGAGAGCCAAAGAAACCCCAAGAAAGCCCGGCTTCAAGTGGACACCCC 422
QY 397 oProArGlySerThrProGluIySgIyProSerThrHisIyProIySgIuIySAlaPheSe 417
DB 423 ACCAAGCTCTCACTCCCAAAAGGAGCCAGAGACCAATCCCAAGAGAGAAAGCTTCTC 482
QY 417 rSerGluIleGluAsPleuProTyIySleuSerThrThrGluMetTyIySleuCyArGTrPhI 437
DB 483 AAGTGAATAGAGATTTGCCGTAACCTTCCACACAGAAATGATTTGGTGGTGGCA 542
QY 437 sGInPProPProSerProIySleuProIySgIuSerSerProIySlySgIuGluThrVa 457
DB 543 CAGAGCTCCCCATCACCGTTACATTACGGGAATCTCTCAAAAGAAAGAGAGAGACTGT 602
QY 457 lAlaArCySleuMePProSerSerValAlaGluThrSerValIySleuAlaValProSe 477
DB 603 AGCA-----AGTAAAGCATAGAGAAACACTTGCTCTT---ATACCTTA 641
QY 477 rTTP 478
DB 642 GTGG 645

RESULT 2
US-08-056-200-93
```

```

; Sequence 93, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054,001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9551 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1507..1644
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1645..2511
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2512..8070
; US-08-056-200-93

Alignment Scores:
Pred. No.: 0.00115 Length: 9551
Score: 238.00 Matches: 162
Percent Similarity: 35.67% Conservative: 82
Best Local Similarity: 23.68% Mismatches: 265
Query Match: 7.35% Indels: 176
DB: 1 Gaps: 30

US-10-054-935-2 (1-614) x US-08-056-200-93 (1-9551)
QY 12 AlaAlaProAlaGluGlyAsnPProGluGluArGLeuAsPryrGluArGAlaAlaLeu 31
DB 4623 GCCGCAAGAGCAGAGAGAGAGAGAGCGGCGCGA-----GAGTGAAGCTGCAATGGCA 4673
QY 32 GlyGlyPProGluAsPgluPProGluIyAlaAlaGluAlaHisPheLeuPro--ArgHisArg 50
DB 4674 GAGAGAGAAAGCGGCTCACCGGCGAGCAGCAGAGAGAGAGAGAGAGCGGCGGACTTCACATG 4733
QY 51 LysLeu-----LysGluPProGlyPro-----Pro 58
```

Db 4734 GCAGTGCAGACCGAGAAAAAGCCAGAGAGGCGCTGTGGCCAGGCCCC 4793
Qy 59 LeuAlaSerSerGlnGly-----GlySerProAlaPro----- 69
Db 4794 ATTCGCGAGACCGGAGAGGAGCAGCTGAGGCGCCAGAGAGCCGACGAGGAGAACAG 4853
Qy 70 ---SerProAlaGlyCysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaGlyAla 88
Db 4854 GTTTCCTCC-----GGAGAGAGGAGAGAGCAGCGCGCCGAGCGAGCGCA 4904
Qy 89 ProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 104
Db 4905 GAGGAG 4964
Qy 105 ProProProAlaThrLysGlnAlaGlyLysGlyGlyGlyGlyGlyGlyGlyGlyGly 122
Db 4965 CCA-----ACAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5015
Qy 123 ---AlaGlyCysSerProArgProLysTyArgLysAlaValLeu----- 135
Db 5016 GCGACAG 5075
Qy 136 -----ProLeginThrGlySerLeuValAlaAlaAlaAlaAlaAlaAlaAla 151
Db 5076 ACGCCGCCACACGCTGTACGCCCAAGCCCTACAGAGAGAGAGAGAGAGAGAGAGAG 5135
Qy 152 TrpAlaGlyAspLysGlyValAlaAlaSerProAlaAlaThrLysSerProAlaGly 171
Db 5136 GCTGCTGCGAG 5192
Qy 172 ProProProLeuProLeuProGlyProPro-----ProLeuAlaProThrAlaGly 190
Db 5193 -----CCAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5240
Qy 191 ThrLeuAlaAlaSerGlnGlyArgTyPlysSerMetArgLysSerProLeuGlyGly 210
Db 5241 GGAG 5294
Qy 211 GlyGlySerGlyLysSerGlnAlaAlaCysLeuLysGlnLysLeuLeuLeuLeu 230
Db 5295 AAGGCAATATCGAGAGAGATTAAGAGCTGCAGCAG--AAGAGAGAGAGAGAGAGAGAG 5352
Qy 231 AspLeuIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 250
Db 5353 GAACCGAG 5412
Qy 251 LysSerGluArgAspThrLeuLeuAlaArgIleGluArgMetGluArgMet----- 268
Db 5413 CAGCAG 5469
Qy 269 -----GlnLeuValLys 272
Db 5470 TGGAG 5529
Qy 273 LysAspAsnGluLysGlnArgHisLysLeuPheGlnGlyTyArgGluArgGlu 292
Db 5530 GAGGAACGGAG 5580
Qy 293 GlnThrGluLeuSerGlnLysLysLysLeuGluCysGlnProGluLeuSerGluThrSer 312
Db 5581 GAG 5628
Qy 313 GlnThrLeuProProLysProPheSerCysGlyArgSerGlyLysGlyHisLysGlyLys 332
Db 5629 GAGACG-----AG 5643
Qy 333 SerProPheGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSer 352
Db 5644 CAG 5697
Qy 353 LysValLysThrLysThrPro-----LysHisSerProIleLysGlnGluProCysGly 370
Db 370 -----LysHisSerProIleLysGlnGluProCysGly 370

Db 5698 CAGCTGCTAG 5757
Qy 371 SerLeuSerGluThrValCysLysArgGluLeuArgSerGlnGluThrProGluLysPro 390
Db 5758 GAG 5817
Qy 391 ArgSerSerValAspThrProProArgLeuSerThrProGlnLysGlyProSerThrHis 410
Db 5818 AGGCGC-----CAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5844
Qy 411 ProLysGluLysAlaPheSerSerGluIleGluAspLeuProTyLysSerThrThrGlu 430
Db 5845 CCGAG 5904
Qy 431 MetTyLysCysArgTrpHisGlnProProProSerProLeuProLeuArgLysSerSer 450
Db 5905 CGAGAGATGTGAATGCGAG 5964
Qy 451 -----ProLysLysGlnGluThrValAlaArgCysLeuMetProSerSerValAla 467
Db 5965 GTTACTGCAAG 6024
Qy 468 GlyGluThrSer-----ValLeuAlaValProSerThrPargAspHis 481
Db 6025 GACAGACAAATCCAGAGAGATGTGCAGACACTGCTGTGAGAGAGAGAGAGAGAGAG 6084
Qy 482 SerValGluProLeuArgAspProAsnProSer-----AspLeuLeu 495
Db 6085 GAGCAG 6144
Qy 496 GluAsnLeuAspAspSerValPheSerLysArgHisLysLysLeu----- 510
Db 6145 GAGCAG 6204
Qy 511 -----GluLeuAspGluLysArgTyLysArgTyLysArgTyLysArgTyLysArg 526
Db 6205 TTGCTGAG 6264
Qy 527 ---GluGlnArgIleLeuGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 541
Db 6265 GAG 6324
Qy 542 LysGlyIleGlnLysSerGlnProGluValThrSerPhePheProGluProAspAspVal 561
Db 6325 AGAAATTCGAG 6345
Qy 562 GluSerLeuMetIleThrProPheLeuProValValAlaPheGlyArgProLeuProLys 581
Db 6345 ----- 6345
Qy 582 LeuThrProGlnAsnPheGluLeuProTyPheLysAspGluArgSerArgGluLeuGlu 601
Db 6346 CTGCTCATAG 6405
Qy 602 IleGlnLysLys 605
Db 6406 CCGAG 6417

RESULT 3
US-08-644-93
Sequence 93, Application US/08800644
Patent No. 5958752
GENERAL INFORMATION:
APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichobyalin and Transglutaminase-3 and
TITLE OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear

```

1 STREET 620 Newport Center Drive, Sixteenth Floor
2 CITY: Newport Beach
3 STATE: CA
4 COUNTRY: U.S.A.
5 ZIP: 92660
6 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Floppy disk
8 COMPUTER: IBM PC compatible
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10 SOFTWARE: PatentIn Release #1.0, Version #1.25
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: US/08/800,644
13 FILING DATE: 14-FEB-1997
14 CLASSIFICATION: 424
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER: US 08/056,200
17 FILING DATE: 30-APR-1993
18 ATTORNEY/AGENT INFORMATION:
19 NAME: Fredrick, Michael F.
20 REGISTRATION NUMBER: 36,799
21 REFERENCE/DOCKET NUMBER: NIH054.001A
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: (714) 760-0404
24 TELEFAX: (714) 760-9502
25 INFORMATION FOR SEQ ID NO: 93:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 9551 base pairs
28 TYPE: nucleic acid
29 STRANDEDNESS: single
30 TOPOLOGY: linear
31 MOLECULE TYPE: cDNA
32 HYPOTHEICAL: NO
33 ANTI-SENSE: NO
34 FEATURE:
35 NAME/KEY: CDS
36 LOCATION: 1507..1644
37 FEATURE:
38 NAME/KEY: intron
39 LOCATION: 1645..2511
40 FEATURE:
41 NAME/KEY: CDS
42 LOCATION: 2512..8070
43 US-08-800-644-93
44
45 Alignment Scores:
46 Pred. No.: 0.00115 Length: 9551
47 Score: 238.00 Matches: 162
48 Percent Similarity: 35.67% Conservative: 82
49 Best Local Similarity: 23.68% Mismatches: 265
50 Query Match: 7.35% Indels: 176
51 DB: Gaps: 30
52
53 US-10-054-935-2 (1-614) x US-08-800-644-93 (1-9551)
54
55 QY 12 A1a1aPProAlaGlyGlyAsnProGluGluArgLeuAspTyrGluArgAla1a1aLeu 31
56 |||||
57 |||||
58 |||||
59 |||||
60 |||||
61 |||||
62 |||||
63 |||||
64 |||||
65 |||||
66 |||||
67 |||||
68 |||||
69 |||||
70 |||||
71 |||||
72 |||||
73 |||||
74 |||||
75 |||||
76 |||||
77 |||||
78 |||||
79 |||||
80 |||||
81 |||||
82 |||||
83 |||||
84 |||||
85 |||||
86 |||||
87 |||||
88 |||||
89 |||||
90 |||||
91 |||||
92 |||||
93 |||||
94 |||||
95 |||||
96 |||||
97 |||||
98 |||||
99 |||||
100 |||||
101 |||||
102 |||||
103 |||||
104 |||||
105 |||||
106 |||||
107 |||||
108 |||||
109 |||||
110 |||||
111 |||||
112 |||||
113 |||||
114 |||||
115 |||||
116 |||||
117 |||||
118 |||||
119 |||||
120 |||||
121 |||||
122 |||||
123 |||||
124 |||||
125 |||||
126 |||||
127 |||||
128 |||||
129 |||||
130 |||||
131 |||||
132 |||||
133 |||||
134 |||||
135 |||||
136 |||||
137 |||||
138 |||||
139 |||||
140 |||||
141 |||||
142 |||||
143 |||||
144 |||||
145 |||||
146 |||||
147 |||||
148 |||||
149 |||||
150 |||||
151 |||||
152 |||||
153 |||||
154 |||||
155 |||||
156 |||||
157 |||||
158 |||||
159 |||||
160 |||||
161 |||||
162 |||||
163 |||||
164 |||||
165 |||||
166 |||||
167 |||||
168 |||||
169 |||||
170 |||||
171 |||||
172 |||||
173 |||||
174 |||||
175 |||||
176 |||||
177 |||||
178 |||||
179 |||||
180 |||||
181 |||||
182 |||||
183 |||||
184 |||||
185 |||||
186 |||||
187 |||||
188 |||||
189 |||||
190 |||||
191 |||||
192 |||||
193 |||||
194 |||||
195 |||||
196 |||||
197 |||||
198 |||||
199 |||||
200 |||||
201 |||||
202 |||||
203 |||||
204 |||||
205 |||||
206 |||||
207 |||||
208 |||||
209 |||||
210 |||||
211 |||||
212 |||||
213 |||||
214 |||||
215 |||||
216 |||||
217 |||||
218 |||||
219 |||||
220 |||||
221 |||||
222 |||||
223 |||||
224 |||||
225 |||||
226 |||||
227 |||||
228 |||||
229 |||||
230 |||||
231 |||||
232 |||||
233 |||||
234 |||||
235 |||||
236 |||||
237 |||||
238 |||||
239 |||||
240 |||||
241 |||||
242 |||||
243 |||||
244 |||||
245 |||||
246 |||||
247 |||||
248 |||||
249 |||||
250 |||||
251 |||||
252 |||||
253 |||||
254 |||||
255 |||||
256 |||||
257 |||||
258 |||||
259 |||||
260 |||||
261 |||||
262 |||||
263 |||||
264 |||||
265 |||||
266 |||||
267 |||||
268 |||||
269 |||||
270 |||||
271 |||||
272 |||||
273 |||||
274 |||||
275 |||||
276 |||||
277 |||||
278 |||||
279 |||||
280 |||||
281 |||||
282 |||||
283 |||||
284 |||||
285 |||||
286 |||||
287 |||||
288 |||||
289 |||||
290 |||||
291 |||||
292 |||||
293 |||||
294 |||||
295 |||||
296 |||||
297 |||||
298 |||||
299 |||||
300 |||||
301 |||||
302 |||||
303 |||||
304 |||||
305 |||||
306 |||||
307 |||||
308 |||||
309 |||||
310 |||||
311 |||||
312 |||||
313 |||||
314 |||||
315 |||||
316 |||||
317 |||||
318 |||||
319 |||||
320 |||||
321 |||||
322 |||||
323 |||||
324 |||||
325 |||||
326 |||||
327 |||||
328 |||||
329 |||||
330 |||||
331 |||||
332 |||||
333 |||||
334 |||||
335 |||||
336 |||||
337 |||||
338 |||||
339 |||||
340 |||||
341 |||||
342 |||||
343 |||||
344 |||||
345 |||||
346 |||||
347 |||||
348 |||||
349 |||||
350 |||||
351 |||||
352 |||||
353 |||||
354 |||||
355 |||||
356 |||||
357 |||||
358 |||||
359 |||||
360 |||||
361 |||||
362 |||||
363 |||||
364 |||||
365 |||||
366 |||||
367 |||||
368 |||||
369 |||||
370 |||||
371 |||||
372 |||||
373 |||||
374 |||||
375 |||||
376 |||||
377 |||||
378 |||||
379 |||||
380 |||||
381 |||||
382 |||||
383 |||||
384 |||||
385 |||||
386 |||||
387 |||||
388 |||||
389 |||||
390 |||||
391 |||||
392 |||||
393 |||||
394 |||||
395 |||||
396 |||||
397 |||||
398 |||||
399 |||||
400 |||||
401 |||||
402 |||||
403 |||||
404 |||||
405 |||||
406 |||||
407 |||||
408 |||||
409 |||||
410 |||||
411 |||||
412 |||||
413 |||||
414 |||||
415 |||||
416 |||||
417 |||||
418 |||||
419 |||||
420 |||||
421 |||||
422 |||||
423 |||||
424 |||||
425 |||||
426 |||||
427 |||||
428 |||||
429 |||||
430 |||||
431 |||||
432 |||||
433 |||||
434 |||||
435 |||||
436 |||||
437 |||||
438 |||||
439 |||||
440 |||||
441 |||||
442 |||||
443 |||||
444 |||||
445 |||||
446 |||||
447 |||||
448 |||||
449 |||||
450 |||||
451 |||||
452 |||||
453 |||||
454 |||||
455 |||||
456 |||||
457 |||||
458 |||||
459 |||||
460 |||||
461 |||||
462 |||||
463 |||||
464 |||||
465 |||||
466 |||||
467 |||||
468 |||||
469 |||||
470 |||||
471 |||||
472 |||||
473 |||||
474 |||||
475 |||||
476 |||||
477 |||||
478 |||||
479 |||||
480 |||||
481 |||||
482 |||||
483 |||||
484 |||||
485 |||||
486 |||||
487 |||||
488 |||||
489 |||||
490 |||||
491 |||||
492 |||||
493 |||||
494 |||||
495 |||||
496 |||||
497 |||||
498 |||||
499 |||||
500 |||||
501 |||||
502 |||||
503 |||||
504 |||||
505 |||||
506 |||||
507 |||||
508 |||||
509 |||||
510 |||||
511 |||||
512 |||||
513 |||||
514 |||||
515 |||||
516 |||||
517 |||||
518 |||||
519 |||||
520 |||||
521 |||||
522 |||||
523 |||||
524 |||||
525 |||||
526 |||||
527 |||||
528 |||||
529 |||||
530 |||||
531 |||||
532 |||||
533 |||||
534 |||||
535 |||||
536 |||||
537 |||||
538 |||||
539 |||||
540 |||||
541 |||||
542 |||||
543 |||||
544 |||||
```

[illegible]

QY 411 ProlyGluLysAlaPheSerSerGluIleGluAspLeuProTyrLeuSerThrThrGlu 430
DB 5845 CGGAGAGAGAGAGAGCTTCCAGCCGACGAGAAAGAGAAACAGGATACCGGATGAGATCAG 5904
QY 431 MetTyrLeuCyArgTrpHisGlnProProProSerProLeuProLeuArgGluSerSer 450
DB 5905 CGCAGTATCTGAAATGCGATGGGAAACGAAAAAGAAATGCAATTCCTGATTAACAG 5964
QY 451 -----ProLysGluGluThrValAlaArgCysLeuMetCProSerSerValAla 467
DB 5965 GTTACTCGCAAGAGAGAGAGATGAACAGTCCGAGAGTTGGAAGATTCACAGTCCG 6024
QY 468 GlyGluThrSer-----ValLeuAlaValProSerThrParGAspHis 481
DB 6025 GACAGCAATCCGACGAAAGATGTCAGACACTGCTGGTGTAACAGAAAGAGATCGT 6084
QY 482 SerValGluProLeuArgAspProAsnProSer-----AspLeuLeu 495
DB 6085 GAGCAAGAGAGAGAGCGCTGCGACAGGCGCAACAGGCAATTTCCAGAGAAACAGCTG 6144
QY 496 GluAsnLeuAspAspSerValPheSerLysArgHisAlaLysLeu----- 510
DB 6145 GAGCGAGAGAGAGAGAAAG 6204
QY 511 -----GluLeuAspGluLysArgArgLysArgTrpAspIleGlnArg 526
DB 6205 TTGCTGAG 6264
QY 527 ---GluGlnArgIleLeuGlnArgLeuGln-----LeuArgMetTyrLysLys 541
DB 6265 GAG 6324
QY 542 LysGlyIleGlnGluSerGluProGluValThrSerPhePheProGluProAspVal 561
DB 6325 AGAAATTCGCGAGAGAGAA----- 6345
QY 562 GluSerLeuMetIleThrProPheLeuProValAlaPheGlyArgProLeuProLys 581
DB 6345 ----- 6345
QY 582 LeuThrProGlnAsnPheGluLeuProTrpLeuAspGluArgCysArgLeuGlu 601
DB 6346 CTGCTCATCAGAGAAACAGGAGAGAAATTCCTCGAGAGAGAGAGAGAGAGAGAGAGAG 6405
QY 602 IleGlnLysLys 605
DB 6406 CGGAGAGAGAA 6417

RESULT 4
US-09-874-923-115
; Sequence 115, Application US/09874923
; Patent No. 6638517
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Meto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, David C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Coler, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; FILE REFERENCE: 210121.420C8
; CURRENT APPLICATION NUMBER: US/09/874,923
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115
; LENGTH: 7065
; TYPE: DNA
; ORGANISM: Leishmania major and chagasi

US-09-874-923-115
Alignment Scores:
Pred. No.: 0.0021
Score: 230.50
Percent Similarity: 36.60%
Best Local Similarity: 24.30%
Query Match: 7.12%
DB: 4
Gaps: 30

US-10-054-935-2 (1-614) x US-09-874-923-115 (1-7065)

QY 10 AlaAlaAlaAlaProAlaGlyGlyAsnProGluGlnArgLeuAspTyrGluArgAlaAla 29
DB 2616 GCTGACAG 2663
QY 30 AlaLeuGlyGlyPro-----GluAspGluProGlyValAlaGluAlaHisPheLeuPro 47
DB 2664 GGCACAGGTGGACAGGCTGGCCGCAACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2714
QY 48 ArgHisArgLysLeuLysGluProGlyProProLeuAlaSerSerGlnGlySerPro 67
DB 2715 AGACACCGCAC----- 2732
QY 68 AlaProSerProAlaGlyCysGlyGly---LysGlyArgGlyLeuLeuProAlaGly 86
DB 2733 GCGGCGCAG 2783
QY 87 AlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 106
DB 2784 GCGGCGCAG 2843
QY 107 ProAlaThrLysGlnAlaGly-----IleGly----- 116
DB 2844 GCAAG 2903
QY 117 GluProAlaAlaAla-----GlyAlaGlyCysSerProArgProLysTyrGln 132
DB 2904 GCGGCGCAG 2963
QY 133 AlaValLeuProIleGlnThrGlySerLeuValAlaAlaAla-----LysGlu 148
DB 2964 GCAAG 3017
QY 149 ProThrProTrpAlaGly-----AspLysGlyGlyAlaAlaSerProAla----- 163
DB 3018 GCGGCGCAG 3077
QY 164 AlaThrAlaSerAspProAlaGlyProProProLeuProLeuProGlyProProProLeu 183
DB 3078 GCAAG 3137
QY 184 AlaProThrAla-----ThrAlaGlyThrLeuAlaAlaSerGlnGlyArgTrpLys 200
DB 3138 GCGGCGCAG 3197
QY 201 SerMetArgLysSerProLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 220
DB 3198 GCAAG 3251
QY 220 AcyLeuLysGlnLe-----LeuLeuLeuGlnLeuAspLeuIleG 234
DB 3252 GCGGCGCAG 3311
QY 234 uGln 247
DB 3312 GCAAG 3371
QY 248 -----GluGluLeuLysSerGluArgAspThrLeuLeuAl 259
DB 3372 GCGGCGCAG 3431
QY 259 ArgIleGluArgMetGluArgMetGlnLeuValLysAsnGlnLysGluArg 279

```

Db      3432 GCAAGCCGCGGAGCTGAGGACGAGGCTGGCGGAGCCGCGGACCGGACGAGCGCGC 3491
Qy      279 ghtslyleu---PheGlnGlyTrpGlnTrpGlnGlnGlnGlnGlnGlnGlnGlnGln 298
Db      3492 CCAGCAGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3551
Qy      298 uhtslyleuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 318
Db      3552 GCGCGCGGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3575
Qy      318 sPhePheSerCysGlyArgSerGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 338
Db      3576 -----GGCAGCGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3622
Qy      338 rGlnArgGlyTrpThrProValIleuLeuAlaProGlnPheSerIleValIleThrIleTrp 358
Db      3623 CGAAGCCGCGAGGAG-----GCTGCAGCAGC 3646
Qy      358 rProIlyshIleSerProIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 376
Db      3647 GCTTACACACCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 3706
Qy      377 -CysIlyshIleArgGlnIleuArgSerGlnGlnIleuTrpProGlnIlyshProArg----- 394
Db      3707 CGGACCGGCGACGAGG--CGCGCCGACGAGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 3765
Qy      394 lAspThrProProArgIleuSerThrProGlnIlyshIlyProSerThrIleProIlyshGly 414
Db      3766 CTAGACACCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 3825
Qy      414 sAlaPheSerSerGlnIleGlnIleuLeuProIlysh----- 426
Db      3826 GACCGCGGAGGAGCGCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 3885
Qy      427 -----SerThrTrpGlnIleuMetIlyleuCysArgTrpIleGlnProProse 442
Db      3886 GACACCGCGACGAGGAGCGCGCGGAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3945
Qy      442 rPro-----LeuProIleuArgGlnIleuSerSerProIlyshGlnGly 455
Db      3946 GCCGAGGAGCTGACGAGCGCTTACACACCGCGACGACGACGACGACGACGACGACGAC 4004
Qy      455 uThrValIleArgCysLeuMetProSerSerValIleGlnIleuThrSerValIleuAla 475
Db      4005 ACGGAGTGGACG-----CT 4019
Qy      475 lProSerTrpArgAspHisSerValGlnProIleuArgAspProAsnProSerAspIleu 495
Db      4020 GCGCGCGGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4076
Qy      495 uGlnIleuLeuAspAspSerValPheSerIlyshArgHisAlaIlyleuGlnIleu----- 512
Db      4077 GCAGCGCTTACACACGCGC-----ACGACGACGCGCGGACGCTGAGGACGAGGAGGAG 4130
Qy      513 -----AspGlnIlyshArgGlnIlyshArgTrpAspIleGlnArgIleu 526
Db      4131 ACGGCTGGCGCGGACCGCGGACGAGCGCGCGACGAG-----CTGCGCGGAG 4178
Qy      526 gGlnGlnArgGlnIleuGlnArgGlnIleuArgMetIlyshIlyshGlnIleGlnI 546
Db      4179 CGCGGAGGAGCTGACGAGCGCTTACACACCGCGACGACGACGACGACGACGACGACGACG 4229
Qy      546 uSerGlnProGlnIleuValThrSerPhePheProGlnProIleuAspAspValGlnSerIleu 566
Db      4230 GCGTGAAGGACAGAGTGGACGCGTGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4289
Qy      566 eThr 567
Db      4290 CACC 4293

```

RESULT 5
US-09-949-016-13141/C

```

; Sequence 13141, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13141
; LENGTH: 20076
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-13141

```

```

Alignment Scores:
Pred. No.: 0.00717 Length: 20076
Score: 230.00 Matches: 138
Percent Similarity: 33.95% Conservative: 45
Best Local Similarity: 25.60% Mismatches: 190
Query Match: 7.10% Indels: 166
DB: 4 Gaps: 25

```

US-10-054-935-2 (1-614) x US-09-949-016-13141 (1-20076)

```

Qy      14 ProIleGlyIleYAsnProGlnGlnIleuAspTrpGlnArgAlaAlaIleuGlyGly 33
Db      2634 CCGGCTGGA-----CCGACACAGCTCTTGCAGCCAAAGAGAGCCCTCTGGGC----- 2584
Qy      34 ProGlnAspGlnProGlnIleuAlaGlnIleuAlaIleuProArgHisArg----- 50
Db      2583 ---AGGAGAGGCTCAGAGGCGGGGTCCAGCCGACGAGACCCCAACGAGGCGCAGGTTT 2527
Qy      51 -----LysIleuIlyshGlnProGlnPro-----Pro 58
Db      2526 GGAACAAAGCGCGCGACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2467
Qy      59 LeuAlaSerSerGlnGlyIleuSerProAlaProSerProAlaGlyCysGlyGlyGly 78
Db      2466 CGCAGTGGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2428
Qy      79 ArgGlyLeuLeuLeuProAlaGlnIleuAlaIleuProGlnGlnGlnGlnGlnGlnGln 98
Db      2427 -----CCGCGGAGAGCTCCGCGAGTCCAGATCCAGACGCGCGCGCGCGCTG 2383
Qy      99 SerValProLeu-----ProCysProProProAlaThrIlyshGlnIleGly 113
Db      2382 AGTCCCCCAATCCGCGCGCGCGGAGCGGTCCAGCGCGCTTTGTTTCCCAAGAGAGC--- 2326
Qy      114 lIleGlyIleuProAlaAlaIleuAlaGlyIleu-----Cys 125
Db      2325 ---CGTGGCGACGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2269
Qy      126 SerProArgPro-IlyshTrpGlnAlaValLeuProIleGlnIleuIlyshSer----- 142
Db      2268 GGTCCAGACCGGAGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2209
Qy      142 uValAlaAlaAlaIlyshGlnIleuProThrProThrProAlaGlnIleuIlyshGlyAlaAlaSer 162
Db      2208 CACGCGTCCAGCGCGGAGGAGTCCGCGCGCTTCCCGGAGGAGTCCAGCGCGCGCGCGCG 2149
Qy      162 oAlaAlaThrAlaSerAspProAla-----GlyProProProProLeuProLeuProGln 179
Db      2148 TCGAGACACCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2092

```

```

QY 179 Y---ProProProLeuAlaProThraAlaThraAlaGlyThrLeuAlaAlaSerGluGlyAr 198
Db 2091 GCGCCCGCCCTCCGCGCCGCGCCGAGCTCCGCGCGCGCGCGCGCGCGCATGCGCCG 2032
QY 198 G---TrrLys---SerMetArg-LyserProLeuGlyGlyGlyGlyGlySerGlyAla 216
Db 2031 GCGCTGCGCGCGCGCGCGCGCTCTCCCTCTCGCGCGCGCGCGCGCGCGCGG----- 1979
QY 216 erserGlnAlaAlaCysLeuLeuGlnLeuLeuLeuGlnLeuAlaSerLeuLeuGlnG 236
Db 1979 ----- 1979
QY 236 lngGlnGlnGlnLeuGlnAlaAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 256
Db 1978 --CAGGTGAGAGATCCCGCGCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGT----- 1925
QY 256 hrLeuLeuAlaAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 275
Db 1924 -----GCGCGGAGAGCACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAAG 1873
QY 276 Gln-----LysGlnArgHisLysLeuPheGlnGlyGlyGlyGlyGlyGlyGlyGly 289
Db 1872 GAGGTGCTGCGCGCGCTTGAGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1813
QY 290 GlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 309
Db 1812 -----AGCTTGCGATGCTCAACAGAGCGCTCCG 1786
QY 310 GlnThrserGlnThrsLeuProGlyProPheSerGlyGlyGlyGlyGlyGlyGlyGly 329
Db 1785 CTCACATCC-----CCCAACACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1744
QY 330 LysArgLysSerProPheGlySerThrsGlnArgLysThrsProValLysLysLeuAlaPro 349
Db 1743 AATTCACGCGCGCGCGCGCGCGCTCGCTCGAG----- 1708
QY 350 GluPheSerLysValLysThrsLysThrsProLysHisSerProLysGlnGlnProCys 369
Db 1707 -----CCCGCGCGAGTCCATCCGCGCGCGCGCTTGC 1675
QY 370 GlySerLeu-----SerGlnThrsValCysLysArgGln----- 380
Db 1674 AACGAGCTCCCTCGGAAAGTGCTCTGCAAAAGCAGCTGCATGAGAGGCGCGCAAG 1615
QY 381 -----LeuArg 382
Db 1614 GCGCAAGGATCGGCGAGCTGCGCGCGCTTCATCTTCATCTTCGCGGAGCTCAT 1555
QY 383 SerGlnGlnThrsProGlnLysProArgSerSerValAspThrsProArgLysSerThrs 402
Db 1554 GCGGCTCAGCGCGCGGATTCCTCAAGAGCGAGGCTGTGCTTGGACATCTCCACT 1495
QY 403 ProGlnLysGlyProSerThrsHisProLysGlnLysAlaLysSerSerGlnLeuGln 422
Db 1494 GCTTGCCCAAGCAGGACCTCACCGAGAGAGTCCCTGCGGTGCGCAAAAGCGCGCA 1435
QY 422 pleuProTyLysSerThrsThrsGlnMetTyLysCysArgTrpHisGlnProProPhe 442
Db 1434 GAGGCTCTGCTGCTCCAGCGATCC-----TGCTCTCTCATCTTCACCTGCAAA 1381
QY 442 rProLeuProLeuArgLysSerSerProLysLysGlnGlnThrsValAlaArgCysLeu 462
Db 1380 CAAGCTCCACCAAGA-----TGCGCGAGATCCAGGTG 1348
QY 462 tProSerSerValAlaGlyGlnThrsSerValLeuAlaValProSerTrp 478
Db 1347 TGGCTTAACAGCTGCGCGCTGAC-----CTACCTAGCTGTCTCATCTGCG 1302

```

RESULT 6
 US-09-949-016-14140/c
 ; Sequence 14140, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:

```

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14140
; LENGTH: 25973
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14140

Alignment Scores:
Pred. No.: 0.00955 Length: 25973
Score: 230.00 Matches: 138
Percent Similarity: 33.95% Conservative: 45
Best Local Similarity: 25.60% Mismatches: 190
Query Match: 7.10% Indels: 166
DB: Gaps: 25

US-10-054-935-2 (1-614) x US-09-949-016-14140 (1-25973)
QY 14 ProAlaGlyLysLysProGlnGlnArgLysLeuAspTyGlnArgAlaAlaLeuGly 33
Db 8531 CCGGCTGGA-----CCCAACAGCTCTCTGACACCAAGAGAGCGCGCTCGGC--- 8481
QY 34 ProGlnAspGlnProGlnAlaAlaGlnAlaHisPheLeuProArgHisArg----- 50
Db 8480 ---AGGAGAGGTACAGGCGGCGGTCCAGCGCGAGAGAGAGAGAGAGAGAGAGTTT 8424
QY 51 -----LysLeuLysGlnProGlyPro-----Pro 58
Db 8423 GAAACAAAGCGCGCGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAGCC 8364
QY 59 LeuAlaSerSerGlnGlnGlySerProAlaProSerProAlaGlyCysGlyGlyLeu 78
Db 8363 CGCAGTGGGATGAGAGAGAGAGCGCGCTGCTCCAGTCC----- 8325
QY 79 ArgGlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGly 98
Db 8324 -----CCCGCGGAGAGCTGCGCGAGATCGACAGATCGACAGCGCGCGCGCTG 8280
QY 99 SerValProLeu-----ProCysProProProAlaThrsLysGlnAlaGly 113
Db 8279 AGTCCCGCATCCCGCGCGCGCGCGCGAGCTGCAAGCGCGCTGTTGTTCCCAAGAGCG--- 8223
QY 114 lLgLyGlyLysProAlaAlaAlaGlyAlaGly-----Cys 125
Db 8222 ---GTTGGGAGACCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8166
QY 126 SerProArgPro-LysTyGlnAlaValLeuProLysGlnThrsLys-----Le 142
Db 8165 GGTCTCAAGCGCGAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 8106
QY 142 vValAlaAlaAlaLysGlnGlnProThrsProTrpAlaGlyAspLysGlyAlaAlaSerPr 162
Db 8105 CACGCGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8046
QY 162 alaAlaThrsAlaSerAlaProAla-----GlyProProProLeuProGln 179
Db 8045 TGCAGAGACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7989
QY 179 Y---ProProProLeuAlaProThraAlaThraAlaGlyThrLeuAlaAlaSerGluGlyAr 198
Db 7988 GCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7929

```

```
QY 198 g----Trrlys---SerMetArg-LysSerProLeuGlyGlyGlySerGlyAla1s 216
Db 7928 GCCCTGACCTGGGACGCGCGCTCTCTCCCTCGCGCGGAGACTGGGCGGG----- 7876
QY 216 exSerGlnAlaIaLysLeuGlyGlnLeuLeuLeuGlnLeuAspLeuIleGluGln 236
Db 7876 ----- 7876
QY 236 lnglnGlnLeuGlnAlaLysGlnLysGlnIleGlnLeuLysSerGluArgAsp 256
Db 7875 --CAGGTGAGATCCCGCGCCCTCCCGCGCCCGCCCTCGCGCGCTGACCGT---- 7822
QY 256 hrLeuLeuAlaIleGlyArg-MetGluArgMetGlnLeuValLysLysAspAsn 275
Db 7821 -----GCGCGGAGGACGAGAGCCCTCGCGGCTCAATTCTCCCGACCTGCAAG 7770
QY 276 Glu-----LysGluArgHisLysLeuPheGlnGlyTyrGlnThrGlu 289
Db 7769 GAGGTGGTGGGCGCTGGGACAGAGGCCCAATCTGTTGGACAGCTGGAAATCAAGA 7710
QY 290 GluArgGlnGlnThrGlnLeuSerGlnLysIleLysLeuGlnCysGlnProGlnLeuSer 309
Db 7709 -----AGCCTCGAGTCAAAAGAGGCTCCCG 7683
QY 310 GluThrSerGlnThrLeuProProLysProPheSerCysGlyArgSerGlyLysGlyHis 329
Db 7682 CTCACCTCC-----CCCAACACCTCGCACAGTCCACACAGCTTGCTGG 7641
QY 330 LysArgLysSerProPheGlySerThrGlnArgLysThrProValLysLysLeuAlaPro 349
Db 7640 AAATTCACGCGCGCGCGGAGCTCGGCTCTGAG----- 7605
QY 350 GluPheSerLysValLysThrLysThrProLysHisSerProIleLysGlnGluProCys 369
Db 7604 -----CCCCGCCAGAGTCCATCCCGCGCGGCTTGC 7572
QY 370 GlySerLeu-----SerGluThrValCysLysArgGlu----- 380
Db 7571 AACCACTCCCTCGGAGAAAGTGGCTCTGCAAAACGACCTGACATGAGAGCGGCACAG 7512
QY 381 -----LeuArg 382
Db 7511 GGGCAAGGATCGGGACGTCGCCCGCTCACCTTCTCCATCTCGGGAGCTCAT 7452
QY 383 SerGlnGlnThrProGlnLysProArgSerSerValAspThrProProArgLeuSerThr 402
Db 7451 GGGCTTCAGCCCGGATCTCTCAAGAGGCGAGGGGTCTGCTGGACCTCCACCT 7392
QY 403 ProGlnLysGlyProSerThr-HisProLysGlnLysAlaPheSerSerGlnIleGluAs 422
Db 7391 GCTTGGCCAAACGACGACCTCACCCAGAGAGTCCCTGGGGTGGCCAAAGCGCGCA 7332
QY 422 pleuProLysLeuSerThrThrGlnMetCysArgTyrPheGlnProProse 442
Db 7331 GAGGCTTCCTGGCTCCCAAGGATCC-----TGCTCTCTCATCTTCCACCTGCAAA 7278
QY 442 rProLeuProLeuArgLysSerSerProLysLysGlnGlnThrValAlaArgCysLeuWe 462
Db 7277 CAGCTGCCACCAAGA-----TGCGCAGATGCGAGTG 7245
QY 462 tProSerSerValAlaGlyGlnThrSerValLeuAlaValProSerTyr 478
Db 7244 TGGCTCTAACAGCTGGGCTGACCT--CTACTAGCTGTCAATCTCG 7199
RESULT 7
US-09-902-540-489/C
; Sequence 489, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
```

```
APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 489
; LENGTH: 2730
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-489
Alignment Scores:
Pred. No.: 0.000936 Length: 2730
Score: 228.50 Matches: 153
Percent Similarity: 33.98% Conservative: 76
Best Local Similarity: 22.70% Mismatches: 237
Query Match: 7.06% Indels: 209
DB: Gaps: 33
US-10-054-935-2 (1-614) x US-09-902-540-489 (1-2730)
QY 14 ProAlaGlyLysAnProGlnGlnArgLeuAspTyrGluArgAlaAlaLeuGlyLys 33
Db 2480 CCCGCGGAGCGCGCGGAA-----TGTTGTCGAGCGGATAGTCGCGGACCTCT 2427
QY 34 ProGluAspGluProGlyAlaAlaGlnAlaHis-----PheLeuProArgHisArgLys 51
Db 2426 CCAACAGGTCCTCTCGGCTCCACGTGACACAGCGCGCGGACCTGTAAGTGGCT 2367
QY 52 LeuLysGlnProGlyProProLeuAlaSerSerGlnGlySerPro--AlaProSer 70
Db 2366 GCCCGCTCCCTCTCTCACCG-----CCTTGAAGCCACAGCG 2331
QY 71 ProAlaGlyCysGlyLys-----Lys 77
Db 2330 CTCGAGATGGCGGAGGACCTCTCTCCGCCACGCGCACGCCACCTGCTCACCC 2271
QY 78 GlyArgGlyLeuLeuLeuPro-----AlaGlyAlaAlaProGlyGlnGlnGlnGlnSer 95
Db 2270 GGGCGAAGCGACGCTCACCTTCCAGAGCGGCTCGCGCGCGGATACAGTCAAGC 2214
QY 96 TrpGlyGlySerValProLeuProCysProProProAlaThrLysGlnAlaGlyLys 115
Db 2213 -----GGCCAGAGGCA---CCGTGCCCCGCGCAGCTCCGAGAGAGCTGCGCGCC 2163
QY 116 GlyLysProAlaAlaAlaGlyLysSerProArgProLysTyrGlnAlaValLeu 135
Db 2162 GGTCAATGGCCGCGACACGCGGTCCATGTCGCCCGCGCTGAGCGGTGATGCT 2103
QY 136 ProIleGlnThrLysLeuValAlaAlaLysGlnLysProThrProThrAlaIleAsp 155
Db 2102 CCGTGT-----GCAAGCCGTGCGCTTTCAGCAAGAGCGGACCTGTG-----AAC 2058
QY 156 LysGlyLysAlaLysSerProAlaAlaThrAlaLysAspProAlaIlePro----- 172
Db 2057 CTGGAAGTTGCGGCACTCAGACACAGCGCGGCTGCTCGTCACTCGCTGTGCG 1998
QY 173 -----ProProLeu 175
Db 1997 CGCGGCCATGACGCGAGGATGCGTGGCGCGGCTTGTCCGAGATGACCGATCAT 1938
QY 176 ProLeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaLys 195
Db 1937 CACGTCACAGAGCGCGGTCTTGCCGTCCAGCGTCTTGAGCTTCTGCGCGCGGTGCGG 1878
QY 196 GluGlyArg----- 198
Db 1877 TACGACACACATCTCTGACCGGACAGCTTCTCAGTGCAGAGGCGTGAAGCGGCTGCC 1818
QY 199 -----TrpLysSerMetArg 203
```



```
Db 1817 CGTACTCCAGGTTTCAGTGTGTCAGTCCACACCGTTGTTGATGCGGCGACCCCG 1758
Qy 204 LysSerProLeuGlyGlyGlyGlyGlyGlyValaSerSerGlnAlaIa-CysLeuLy 223
Db 1757 ---CGGCTTCAGCGCGGTCTCTGATCCATCCAGGAGCGGCGCATCTTCAGTTCTTCCA 1701
Qy 223 sgnIle-----LeuLeuLeuGlnLeuAspLeuI 233
Db 1700 CCAAGCGCGCGCAGTACGCGGCGGCGGTCGCGGCTCCAGCGGCGACCTTCACCTTGCT 1641
Qy 233 eGluGlnGlnGlnGlnGlnGlnGlnAlaLysGlnLysGlnLysGlnLysSerG 253
Db 1640 CGGCGCGGCGGCTGCGGCTCCGAGCTCCGCGGCTTCGCGGCGGCGGCGGCGGCGG 1581
Qy 253 uArgSerThrLeuLeuAlaArgIleGluArgMetGluArgMetGlnLeuValLysLy 273
Db 1580 TCACCAACGCTCAGCTCCGCGGCGGCGGCGGCGGCGGCTGAGCGGCTCCGCGGCTTCGCGG 1521
Qy 273 sAspAsnGlnLysGlnLysGlnLysLeu-----PheGlnGlnLysGlnLysGln 289
Db 1520 TGAAGTTACAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1461
Qy 289 uGluArgGluGluThrGln-LeuSerGlnLysIleLysLeuGlnCysGlnProGlnLeuS 309
Db 1460 CGGCGGCTGCGGCGGAGAGATGAGCAGCGCGCTGAGCTCCGCTCAGCGGCGGCGGCGG 1404
Qy 309 eGlnLysSerGlnThrLeuPro-----ProLysProPheSerCysGlnLysArgS 325
Db 1403 CTTTGACGAGCAGCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1344
Qy 325 eGlnLysGlnLysLysLysLysSerProPheGlnSer-ThrGluArgLysThrProVal 344
Db 1343 CC-----ATGCCGTTGGCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1305
Qy 345 LysLysLeuAlaProGlnPheSerLysValLysThrLysThrProLysHisSerProIle 364
Db 1304 GCACCTTGTCGCGGCGGAGTTCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1245
Qy 365 LysGlnGluProCysGlnSerLeuSerGlnThrValLys----- 377
Db 1244 CGGCGA-----TGTCCAGCTGCGGTCGAGCGGAGCTTGTCCGCGGCGGCTGTCAGCGG 1191
Qy 378 -----LysArgGln-----LeuArgSerGlnLysThrPro 387
Db 1190 ACTGCTGATTTGGCCACCAACCGCGCGGAGGCTTCGCGGCGGCGGCTTCATTCCT 1131
Qy 388 GlnLys-----ProArgSerSerValAspThrProProArgLeuSerThr----- 402
Db 1130 CAATCTCCAGGCGCGGCGGCTGAGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1071
Qy 403 ProGlnLysGlnProSerThrHisProLysGlnLysAlaPheSerSerGlnLysGlnAsp 422
Db 1070 CCAAGTATGCGCGGCGGCTTCACCGAAA-----TCTTCACAGGTCACACCT 1023
Qy 423 Leu-----ProTyrLeuSerThrThrGlnMetCysThrLysCys 434
Db 1022 CTTTCATCATCGGCTGAGAACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 963
Qy 435 ArgTyrHisGln----- 438
Db 962 TCAGAGAACCGGCGGCTGCTTCGAAACATCATCCGAGGTGTCGATCGGCTAGCGGAGC 903
Qy 439 -----ProProProSerProLeuProLeuArgLys-----SerSer 450
Db 902 ATGGCGAGTGGCTCCAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 843
Qy 451 ProLysLysGlnGluThrValAlaArgCysLeuMetProSerSerValAlaGlnLysThr 470
Db 842 CGGCGGAGGTGAAGAGCTTGGGG---TGCACCATGCG----- 807
Qy 471 SerValLeuAlaValProSerThrArgAspHisSerVal---GluProLeuArgAspPro 489
```

```
Db 806 -----CTGCCAGACCTCCACACCGGCTGATTCAGATGCGGAGCCC 759
Qy 490 AsnProSerAspLeuLeuGln-----AsnLeuAspAspSerVal----- 502
Db 758 TTGCCCGCGGAGTTCCGAGGAGGTGATGATCCACTCGGCGGAGCGGCTCCGTGAAGGAG 699
Qy 503 -----PheSerLysArgHisAlaLysLeuGlnLeuAspGlnLysArgLysArgTyr 520
Db 698 AAGGACTGCGGGAAGCGGCT-GCGCGTGTCCGAACGGAAGACGCGCTCAGCAA----- 646
Qy 521 AspIleGlnArgIleArgGlnGlnArgIleLeuGln 532
Db 645 -----CGCGCGCAACGAGCGCTTCAG 625

RESULT 8
US-09-620-312D-713
Sequence 713, Application US/09620312D
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aiding J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungang
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhixue
APPLICANT: John Tillinghaast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 656962e1 Nucleic Acids and
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc FL_genes Version 1.0
SEQ ID NO 713
LENGTH: 3169
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (116)..(2146)
US-09-620-312D-713

Alignment Scores:
Pred. No.: 0.00111 Length: 3169
Score: 228.50 Matches: 151
Percent Similarity: 36.25% Conservative: 77
Best Local Similarity: 24.01% Mismatches: 199
Query Match: 7.06% Indels: 204
DB: 4 Gaps: 32

US-10-054-935-2 (1-614) x US-09-620-312D-713 (1-3169)
Qy 13 AlaProAlaGlyGlyLysProGlnGlnArgLeuAspTyrGlnArgAlaAlaLeuGly 32
Db 47 AGTCCGCGCTCCGCGGCTCT----- 67
Qy 33 GlyProGlnAspGlnProGlnAlaAlaGlnAlaHisPheLeuProArgHisValGlyLeu 52
Db 68 TCACACATGATCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 127
```

```

QY 53 LysGluProGlyProProLeuAlaSerSerGlnGlySerProAlaProSerProAla 72
DB 128 AAACAAAGGGGCG-----TCACAGGGGGCCACGCGCCGCGAGACCCCTCG 172
QY 73 GlyCysGly-----GlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaIlePro 89
DB 173 GAGAGGGGTGCCACCCGCTGCGCGCGCGCTCTCTGCGCTCCAGAGCCAGAGAGGCC 232
QY 90 GlnGlnGlnGlnGlnSerTrpGlyGlySerValProLeuProCysProProAlaThr 109
DB 233 GCGTCCGCGTCCGCGCAGTCCCGGGGGAGCC-----CTGCAGATTCCCGGGCGCG 283
QY 110 LysGlnAlaGlyIleGlyGlyProAlaAlaAlaGlyAlaGlyCysSerProAlaGPro 129
DB 284 GCGCCGCGCTGCTGCTGCTCCAGCCCGCGCGCGAGCCGCGCGCGCGCGCATGCC 343
QY 130 LysTyrGlnAlaValLeuProIleGlnThrGlySerLeuValAlaAlaIleLysGluPro 149
DB 344 -----TCGGCCAAACAA----- 355
QY 150 ThrProTrpAlaGlyAspLysGly-----GlyAlaAlaSerProAlaAlaThrAlaSer 167
DB 356 -----AGGGGCTCCAAAGGCGGCGCAAGCGCGCGAGCCCTCTCGAGAAAGGGTCC 406
QY 168 AspProAlaGly-----ProProProLeuProProLeuProGly 179
DB 407 CACCCCTCGGCGCGCGATGACGTGGCGAGAAAGCGCGCGCGCGCGCGAGCGCG 466
QY 180 ProProProLeuAlaPro-ThrAlaThrAlaGlyThrLeuAlaAlaSerGlnIleArgTr 199
DB 467 CCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
QY 199 LysSerMetArgLysSerProLeuGlyGlyGlyGlySerGly----- 214
DB 524 GCGCAGGCGCAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 582
QY 215 -----AlaSerSerGlnAlaAlaCysLeuIle 223
DB 583 CTCCTCCGCTCCGCGCGCGCTGCGCGCGCGCGCGCGCTCTCTCGCGCTCGCG 642
QY 223 GlnIle-----LeuLeuLeuGlnLeuAspLeuIle----- 233
DB 643 CAGGCTCGGAGGCGCTCACTTCTCTTCACTCGCGCTGCTGCGCGCGCGCGCTTT 702
QY 233 ----- 233
DB 703 CTCGGGCTGGTGGCTCCACCAAGCTCGAGAGGTCACAGGTCGCGCGCGAGCCCA 762
QY 234 -----GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 247
DB 763 GGAATTCTCCCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 822
QY 247 GlnGlnGlnLysSerGlnLysArgSerThrLeuLeuAlaArgIleGlnLysMetGlnArg 267
DB 823 GCGCTTTTGCAAGCCCATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 882
QY 267 GlnGlnLysValLys-----AspAsnGlnLysGlnArg-----HisIle 281
DB 883 ACAAGACCTCACAGAGAAAGCTGTGAAGCAAGGGGAGAGTGAAGTCAAGCCGATCAGCG 942
QY 281 sLeuPheGlnIleTyrGlnIleThrGlu----- 289
DB 943 AGTGCTGAGAAACTCCAGAAATGATTTCAAAAGACCTCTCGAGTGGATTCATGTGT 1002
QY 290 -----GlnArgGlnIleThrGlnLysSerGlnLysIleLysLeuGlnCysGln 305
DB 1003 GAAGACGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1059
QY 305 nProGlnLeuSerGlnThr-SerGlnThrLeuProProLysProPheSerCysGlyArg 325
DB 1060 GACGAGCTCACCAATTCATCAACAGACACATCGCCA-----TCTTCACAGAGT 1110

```

```

QY 325 erGlyLysGlyHisLysArgLysSer-----ProPhe----- 335
DB 1111 CCAGAGAGAGGAGCCAGAGAGAGATCATATGATCAATGAAGCAAGTTGCTCCTCGGAGAA 1170
QY 336 -----GlySerThrGlnArgLysThrProValLysLysLeu----- 347
DB 1171 ACTGAGGGGAACAGACAGAGATTGAAACCTTTAAAGAAAGCTGTGAAGAGATACAGAC 1230
QY 348 -----AlaProGlnPheSerLysValLys----- 355
DB 1231 CTCAGCCAGATCCAGAGATGGAGCATGAGAGCCCTGAGAAATACCTTCAGACTATGGA 1290
QY 356 -----ThrLysThrProLysHisSer----- 362
DB 1291 GTCGTACATCTACCCGAGATCCGCGAGCTGTGATGACCTCAACAGACAGACAGGCTTT 1350
QY 363 -----ProLysGln-----GlnProCysGlySerLeuSerGlnThrValCysL 378
DB 1351 CAAGAGGCGCGCGACACAGAGAGCGCTGCGCTGCGAG-GCCTCAGAGAAAGCTTCTCA 1409
QY 378 LysArgGlnLeuArgSerGlnIleThrProGlnLysProArgSerSerValAspThrProp 398
DB 1410 GGTCTGAGGAGTCCGCTCCGCGCTCCGAGAGATCCGAGATCGAGAGAAAGCTCC 1469
QY 398 roArgLysSerThrProGlnLysGlyProSer-----ThrIleProLysG 413
DB 1470 GCGAGCTGAAGTCCGATTCACGCGCGCGAGAGAGAGAGAGGCTTCAGACACTCGAAG 1529
QY 413 LysAlaPhe-SerSerGlnIleGlnAsp-----LeuProTyrLeuSerThrThrGln 430
DB 1530 CCTTGAAGCATCTCCGCAAAAGATCAGAGACTGATCTCCAGCTCCAGACGTGGAAG 1589
QY 431 MetTyrLeu-----CysArgTrp-----HisGlnPro-ProProSerProLeuPr 445
DB 1590 ATGGGTGTCTTCATGACGATGAGTGTCTGCGCGCGACGAGACCTGAGATCCCTCC 1649
QY 445 oleuArgGlnLysSerProLysLysGlnIleThrValAlaArgCysLeuMetProSerSe 465
DB 1650 TGTCCAAAG-----CCAGAGACAGAGACGCGCTGCGCG-----CTTCAGAG 1694
QY 465 rValAlaGlyIleThrSerValLeu-----AlaValProSerTrp--ArgAspHisSerVa 483
DB 1695 GCGGCTGGAAGCGCTCGGCTCTTCAGAGAGACAGACAGATGGCTGCGCAGCAGGTGA 1754
QY 483 LgluProLeuArgAspPro 489
DB 1755 GAGGCTGGGCGAGACCA 1773

RESULT 9
US-09-902-540-1096
; Sequence 1096, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1096
; LENGTH: 14555
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(14555)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1096

```



```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL0010307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12498
; LENGTH: 5788
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-12498

Alignment Scores:
Pred. No.: 0.00379      Length: 5788
Score: 224.00          Matches: 151
Percent Similarity: 36.36%      Conservative: 53
Best Local Similarity: 26.92%      Mismatches: 220
Query Match: 6.92%          Indels: 138
DB: 4                  Gaps: 29

US-10-054-935-2 (1-614) x US-09-949-016-12498 (1-5788)

QY 10 AlaAlaAlaAlaPro-----AlaGlyGlyAsnProGluGlnArgLeuAsp 24
Db 2543 GCAGCTGAGAGCTCCAGAGCAGAGAGGAGGAGCTGAGCTCC-----2587
QY 25 TyrGluArgAlaAla-----AlaLeuGlyGlyProGluAspGluProGlyAlaAlaGlu 42
Db 2588 -----AGAGCAGCAGAGAGGAGGAGCTGAGCTCCAGAGCAGCAGAGGAGGAGCTGGA 2641
QY 43 AlaHisPheLeuProArgHisArgLysLysLeuLysGluProGlyProProLeuAlaSerSer 62
Db 2642 GCT-----CCAGAGCAGCAGAGAGGAGGAGCTGAGCTCCAGAGCAGCAGAGAGG 2692
QY 63 GlnGlyGlySerPro-----AlaProSerProAlaGlyCysGlyGly 76
Db 2693 GCAGCTGAGAGCTCCAGAGCAGCAGAGAGGAGGAGCTGAGCTCCAGAGCAGCAGAGAGG 2752
QY 77 LysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnSerTyr 96
Db 2753 ---GCAGCTGAGAGCTCTGAGCAGCAGAGGAGGAGCTGAGCTCTGAGCAGCA---2806
QY 97 GlyGlySerValProLeuProCysProPro-----106
Db 2807 GAGGAGCAGCTGAGAGCAGCTGAGCAGCAGAGGAGGAGCTGAGCTCCAGAGAGCA 2866
QY 107 -----ProAlaThrLysGlnAlaGlyTyrLeuGlyGluProAlaAla 120
Db 2867 GATGGGCGACCTGAAGTACTGGAACA---GCAGAGAGGAGCAGTGAAGCAGCTGATCA 2923
QY 121 AlaGly---AlaGlyCysSerProArgProLysTyrGlnAlaVal---LeuProIleGln 138
Db 2924 GCAGGAGAGCAGCAGAGAGCTCCAGAGCAGAGATGGAGGAGCAGTGAACCTGGAGCA 2983
QY 139 ThrGlySerLeuVal-----AlaAlaAlaLysGluProThrProTyrAlaGlyAsp 155
Db 2984 GCAGAGAGGAGCAGCTTAAGCATCTGAGCAGCAGAGAGGAGCAACTGAGAGCTGGAGGA 3043
QY 156 LysGlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeu 175
Db 3044 GCAGAGAGGAGCAGCTGAAGCAGCTGAGAGCAGCAGAGGAGGAGCAGCTGAGCACT-----3097
QY 176 ProLeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSer 195
Db 3098 GAGACACAGAGAGAGGAGCAGCTGGGGCTCCAGAGCA---GCAAGTGTGTGAGCTGAAGCA 3154
QY 196 GlnGlyArgTyrLysSerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAla 215

```

```

Db 3155 GCTAGAGAGCAGCAGAGGAGCAGCAAGACCTTGAGAGAGAGAGGAGGAGCTGAAGCA 3214
QY 216 SerSerGlnAlaAlaCysLeuLysGlnIle-LeuLeuLeuGlnLeuAspLeuIleGluGlu 235
Db 3215 CCTGTGTCAGCAGAGAGGAGCAGCTGAAGCATCTGTGTGTCAGCAGAGAGGAGCAGCTGAGCA 3274
QY 235 nGlnGlnGlnGlnGlnGlnAlaLysGluLysGluIleGluGluLeuLys-----SerGlu 253
Db 3275 GCAGAGAGCAGCTGTGAAGCAGCTGAGCAGAGAGGAGGAGCTGAACCACTTAAGAGA 3334
QY 253 uArgAspThrLeuLeuAlaArgIleGluArgMetGluArgAspMetGlnLeuValLys--272
Db 3335 GCAGAGAGCAGCACTAAGCATCTGAGAGCAGCAGAGGAGGAGCTGAGAGCTCCAGAGCA 3394
QY 273 -----LysAspAsnGluLysGluArgHisLysLysLeuPheGlnIleTyrGlu 287
Db 3395 GCAGGTGGGAGCAGCAAGAAAGACCTGAGAGCAGAGAGAGCAACTG-----GA 3442
QY 287 uThrGluGluArgGluGluGluThrGluLeuSerGluLysIleLysLeuGluCysGlnProGlu 307
Db 3443 GCTCCAGAGCAGCAAGAGAGGAGGAGCTGAAGCAGCTGAAGAGCAGAGCAGCACTGGA 3502
QY 307 uLeuSerGluThrSerGlnThrLeuProLysProPheSerCysGlyArgSerGlyTyr 327
Db 3503 GCTCCAGAGCAGCAGCTGAGAGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 3552
QY 327 sGlyHisLysArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLe 347
Db 3553 A-----3553
QY 347 uAlaProGluPheSerLysValLysThrLysThrPro-LysHisSerProIleLysGluGlu 367
Db 3554 -GCACCTTAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3612
QY 367 LuProCysGlySerLeuSerGluThrValCysLysArgGluLeuArgSerGlnGluThrP 387
Db 3613 GGCAGCTGAAGACCTTGAGAGCAGAG-----AAGGGAGAGCTGAGAGCAGCTGTGTTG 3666
QY 387 roGluLysProArgSerSerValAspThrProProArgLeuSerThrProGlnLysGly-406
Db 3667 CCCAGACTCAGGCCAGGCTCAGGCTCAAGACATTCACACACCTG-----CCACAAAGAGAG 3720
QY 407 -----ProSerThrHisProLysGluLysAlaPheSerSerGluIleGluAspL 423
Db 3721 AAGTATTGCTTCTCTGAGAGCAGCAGCAGCAGAG-----CAGAGGTGTCAG----3767
QY 423 euProTyrLeuSerThrThrGlnMetTyrLeuCysArgTyr-His-----Gln 438
Db 3768 -----TGGCCACCCCAACATTAATTAAC 3789
QY 439 ProProProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrValAla 458
Db 3790 CACCCAGAGCTCAGAGAGGAGCAGCTGAGTGTCTCATCAAGAGAGAGAGAGCAGCAGTGC 3849
QY 459 ArgCysLeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSer---477
Db 3850 TCCACTTATTGCGGCTCCCTAGGTGCGCCGTCATCTGTGAACCTTGACTGTGCTCTC 3909
QY 478 -----TyrArgAsp-----HisSerValGlu 484
Db 3910 TACATGTCTCTTAAATGAGGAGTGAAGGTGGAGAGAGAGAGAAATTATGTTCAGGTGCAAC 3969
QY 485 ProLeuArgAspProAsnProSer-----AspLeuLeuGlnLeuLeu 498
Db 3970 CCC---AATGACCCCAATCCCAACCTCAAGTGAAGCATGGCCCTCTATCTTAAGGAGACTA 4023

RESULT 11
US-09-949-016-14458
; Sequence 14458, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

```

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: C1001307
 CURRENT APPLICATION NUMBER: US/09/949,016
 PRIOR FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 14458
 LENGTH: 5788
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-14458

Alignment Scores:
 Pred. No.: 0.00379
 Score: 224.00
 Percent Similarity: 36.36%
 Best Local Similarity: 26.92%
 Query Match: 4
 Matches: 5788
 Conservative: 151
 Mismatches: 53
 Indels: 220
 Gaps: 138
 29

US-10-054-935-2 (1-614) x US-09-949-016-14458 (1-5788)

QY 10 AlaAlaAlaAlaPro-----AlaGlyGlyAsnProGlnGlnArgLeuAsp 24
 Db 2543 GCAGCTGAGAGCTCCAGAGCAGAGGAGGAGGAGCTGAGCTCC----- 2587
 QY 25 TyGlnArgAlaAla-----AlaLeuGlyGlyProGlnAspGlnProGlnAlaAlaGlu 42
 Db 2588 -----AGAGCAGCAGAGGAGGAGGAGCTGAGCTCCAGAGCAGAGGAGGAGCTGGA 2641
 QY 43 AlaHisPheLeuProArgHisArgGlyLeuGlyGlnProGlnProLeuAlaSerSer 62
 Db 2642 GCT-----CCAGAGCAGCAGAGGAGGAGGAGCTGAGCTCCAGAGCAGAGGAGG 2692
 QY 63 GlnGlyGlySerPro-----AlaProSerProAlaGlyGlyGlyGly 76
 Db 2693 GCAGCTGAGAGCTCCAGAGCAGAGGAGGAGGAGCTGAGCTCCAGAGCAGAGGAGG 2752
 QY 77 LysGlnArgGlyLeuLeuLeuProAlaGlyAlaAlaProGlnGlnGlnGlnGlnSerTyr 96
 Db 2753 ---GCACTGAGAGCTCTGAGAGCAGAGGAGGAGCTGAGCTCTGAGAGCA--- 2806
 QY 97 GlnGlySerValProLeuProCysProPro----- 106
 Db 2807 GAGAGGACAGCTGAGAGCAGAGCAGAGGAGGAGCTGAGCTCCAGAGAGCA 2866
 QY 107 -----ProAlaThrLysGlnAlaGlyTyrLeuGlyGlnProAlaAla 120
 Db 2867 GATGGGAGAGCTGAGAGTCTGAGAGCA---GAGAGAGGAGCAGTGAAGAGCAGTGA 2923
 QY 121 AlaGly---AlaGlyCysSerProArgProLysTyrGlnAlaVal---LeuProLysGln 138
 Db 2924 GCAGGAGAGAGCAGAGCTCCAGAGCAGAGTGGAGGAGCTGAGAGCACTGAGAGCA 2983
 QY 139 ThrGlnSerLeuVal-----AlaAlaAlaLysGlnProThrProTyrAlaGlyAsp 155
 Db 2984 GCAGAGAGGAGCAGCTGAAGATCTGAGAGCAGAGGAGGAGCAGTGAAGAGAGCA 3043
 QY 156 LysGlnGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProLeu 175
 Db 3044 GCAGAGAGGAGCAGCTGAAGAGCTGAGAGCAGAGAGGAGGAGCTGAGAGCACT--- 3097
 QY 176 ProLeuProGlnProProProLeuAlaProThrAlaThrAlaThrLeuAlaAlaSer 195
 Db 3098 GAGAGCAGAGAGAGGAGAGCTGGGCTCCAGAGCA---GAGGTGCTGAGCTGAGAGCA 3154

QY 196 GlnGlyArgTyrIleSerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAla 215
 Db 3155 GCTAGAGAGCAGCAGAGGAGGAGCAGCAAGAGCTGAGAGAGAGAGGAGGAGCTGAAGCA 3214
 QY 216 SerSerGlnAlaAlaCysLeuLysGlnIle-LeuLeuLeuGlnLeuAspLeuLeuGlu 235
 Db 3215 CTTGTGAGAGCAGAGGAGGAGCTGAGAGCAGTGTGAGAGCAGAGGAGGAGCTGAGAGCA 3274
 QY 235 ngInGlnGlnLeuGlnAlaLysGlnLysGlnIleGlnGlnLeuLys-----SerGln 253
 Db 3275 GCAGAGAGGAGCAGTGAAGATCTGAGAGCAGAGGAGGAGCTGAGAGCACTGAGAGCA 3334
 QY 253 uArgAspThrLeuLeuAlaArgIleGlnArgMetGlnArgArgMetGlnLeuValLys-- 272
 Db 3335 GCAGAGAGGAGCAGTGAAGATCTGAGAGCAGAGGAGGAGCAGTGAAGAGCA 3394
 QY 273 -----LysAspAsnGlnLysGlnLysGlnLysLysLeuPheGlnGlnLysGln 287
 Db 3395 GCAGTGGAGGAGCAGCAAGAGCCTGAGAGCAGAGAGAGAGCAAGCACTG---GA 3442
 QY 287 uThrGlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 307
 Db 3443 GCTCCAGAGCAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3502
 QY 307 uLeuSerGlnThrSerGlnThrLeuProProLysProPheSerCysGlnArgSerGly 327
 Db 3503 GCTCCAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3552
 QY 327 sGlnHisLysArgLysSerProPheGlnSerThrGlnArgLysThrProValLysLys 347
 Db 3553 A----- 3553
 QY 347 uAlaProGlnPheSerLysValLysThrLysThrPro-LysHisSerProLysGlnGlu 367
 Db 3554 -GCAGCTAGAGCAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGG 3612
 QY 367 LysProGlnSerLysSerLysSerLysThrValCysLysArgLysLysLysSerGlnGlu 387
 Db 3613 GGCAGCTGAAGAGCCTGAGAGCAGAG-----AAGGGAGAGCTGAGAGCAGCTGTGTTG 3666
 QY 387 roGlnLysProArgSerSerValLysThrProProArgLeuSerThrProGlnLysGly 406
 Db 3667 CCCAGCTCAGAGCAGAGCTCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGG 3720
 QY 407 -----ProSerThrHisProLysGlnLysAlaPheSerSerGlnLysGlnLys 423
 Db 3721 AAGTATTCTTCTGTAGAGCAGAGCAGAGAG-----CAGAGGTGAG--- 3767
 QY 423 euProTyrLeuSerThrThrGlnMetTyrLeuCysArgTyrHis-----Gln 438
 Db 3768 -----TGGCAGCCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAG 3789
 QY 439 ProProProSerProLeuProLeuArgLysSerProLysLysGlnGlnThrAlaAla 458
 Db 3790 CACCCGAGAGTGTGAGAGGAGCAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3849
 QY 459 ArgCysLeuMetProSerSerValAlaGlyLysThrSerValLeuAlaValProSer--- 477
 Db 3850 TCCACTTATTTCCGGTCCGCTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3909
 QY 478 -----ThrArgAsp-----HisSerValGlu 484
 Db 3910 TACATGCTCTTTAATAGGGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3969
 QY 485 ProLeuArgAspProAsnProSer-----AspLeuLeuGlnLeuLeu 498
 Db 3970 CCC---AATGACCCCAATCCCAACTCAGGTGAGAGCAGGCTTACTTGTGAGGAGCTA 4023

RESULT 12
 US-09-949-016-15247/c
 ; Sequence 15247, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 15247
LENGTH: 5870
TYPE: DNA
ORGANISM: Human
US-09-949-016-15247

Alignment Scores:
Pred. No.: 0.00385 Length: 5870
Score: 224.00 Matches: 151
Percent Similarity: 36.36% Conservative: 53
Best Local Similarity: 26.92% Mismatches: 220
Query Match: 6.92% Indels: 138
Gaps: 29
DB:

US-10-054-935-2 (1-614) x US-09-949-016-15247 (1-5870)

QY 10 AlaAlaAlaAlaPro-----AlaGlyGlyAsnProGluGlnArgLeuAsp 24
DB 2934 GCAGCTGAGAGCTCCAGAGCAGAGAGGGGAGCTGAGCTCC----- 2890

QY 25 TyGluArgAlaAla-----AlaLeuGlyGlyProGluAspGluProGlyAlaAlaGlu 42
DB 2889 -AGAGCAGCAGAGAGGGGAGCTGAGCTCCAGAGCAGCAGAGAGGGGAGCTGCA 2836

QY 43 AlaHisPheLeuProArgHisArgLysLeuLysGluProGlyProLeuAlaSerSer 62
DB 2835 GCT-----CCAGAGCAGCAGAGAGGGGAGCTGAGCTCCAGAGCAGCAGAGAGG 2785

QY 63 GlnGlyGlySerPro-----AlaProSerProAlaGlyGlyGlyGly 76
DB 2784 GCAGCTGAGAGCTCCAGAGCAGAGAGGGGAGCTGAGCTCCAGAGCAGAGAGAGG 2725

QY 77 LysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGln 96
DB 2724 -GCAGCTGAGAGCTCTCTGAGAGCAGAGAGGGGAGCTGAGCTCTCTGAGAGCA--- 2671

QY 97 GlyGlySerValProLeuProCysProPro----- 106
DB 2670 GGAGGAGCAGCTGAGAGCAGCTGAGAGCAGAGAGGGGAGCTGAGAGTCCAGAGAGCA 2611

QY 107 -----ProAlaThrLysGlnAlaGlyLysGlyGlyGluProAlaAla 120
DB 2610 GATGGGAGAGCTGAGAGTACCTGAGACA---GCAGAGGGGAGAGCTGAGAGCTGAGTCA 2554

QY 121 AlaGly-----AlaGlyCysSerProArgProLysTyGlnAlaVal---LeuProIleGln 138
DB 2553 GCAGAGAGAGCAGCAGAGAGCTCCAGAGCAGAGAGTGGGGAGCTGAGAGCACTGAGCA 2494

QY 139 ThrGlySerLeuVal-----AlaAlaAlaLysGluProThrProThrProAlaGlyAsp 155
DB 2493 GCAGAGAGGGGAGAGCTTAAGCATCTGAGAGCAGAGAGGGGAGCACTGAGAGAGTGA 2434

QY 156 LysGlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProLeu 175
DB 2433 GCAGAGAGGGGAGAGCTGAGAGCACTGAGAGCAGAGAGGGGAGCACTGAGAGCACT--- 2380

QY 176 ProLeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSer 195
DB 2379 GAGAGCAGAGAGAGGAGAGCTGGGGCTCCAGAGCA---GCAGGTGCTGAGCTGAGAGCA 2323

QY 196 GluGlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGlyGlyGlyGlyAla 215
DB 2322 GCTAGAGAGCAGCAGAGGGGAGAGCTGAGAGAGAGAGAGGGGAGAGCTGAGAGCA 2263

QY 216 SerSerGlnAlaAlaCysLeuLysGlnIle-LeuLeuGlnLeuAlaSerLysGlnIle 235
DB 2262 CCTGTGTCAGCAGAGAGGGGAGAGCTGAGAGTCTGTGTCAGCAGAGAGGGGAGAGCTGAGAGCA 2203

QY 235 ngGlnGlnGlnGlnLeuGlnAlaLysGluLysGluLysGluLysGluLys-----SerGln 253
DB 2202 GCAGAGAGAGAGCTGAGAGCACTGAGAGCAGAGGTGGGGAGCTGAGAGCACTGAGAGCA 2143

QY 253 ValArgAspThrLeuLeuAlaArgLysGluArgMetGluArgMetGlnLeuValLys--- 272
DB 2142 GCAGAGAGGAGCAACTGAGACATCTGAGAGCAGAGAGGGGAGAGTGGAGTCCAGAGCA 2083

QY 273 -----LysAspAsnGluLysGluArgHisLysLysLeuPheGlnGlyTyrGln 287
DB 2082 GCAGGTGGGAGAGCCAAAGAACCTGAGAGAGAGAGAGCAACTG-----GA 2035

QY 287 uThrGlnGluArgGluGluGluThrGluLeuSerGluLysLysLysLeuGluCysGlnProGln 307
DB 2034 GCTCCAGAGCAGCAAGAGAGGGGAGAGTGAAGACCTGAGAGAGAGAGAGAGAGAGAGCTGAG 1975

QY 307 uLeuSerGluThrSerGlnThrLeuProProLysProPheSerCysGlyArgSerGlyLys 327
DB 1974 GCTCCAGAGCAGCAGAGTGAAGACAGCCAA--GCAGCT-----GGAACAGCAGAGAAA 1925

QY 327 eGlyHisLysArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLe 347
DB 1924 A----- 1924

QY 347 uAlaProGluPheSerLysValLysThrLysThrPro-LysHisSerProIleLysGluG 367
DB 1923 -GCAGCTTAAGAGCAGCAGAGAGAGAGAGAGAGCACTGAACATCTGAGAGAGAGAGAG 1865

QY 367 LuProCysGlySerLeuSerGluThrValCysLysArgGluLeuArgSerGlnGluThrP 387
DB 1864 GGCAGCTGAAGAGCCTGAGAGCAGAG-----AAGGGGAGCTGAGAGAGAGCTGTGTTG 1811

QY 387 roGluLysProArgSerSerValAspThrProProArgLeuSerThrProGlnLysGly- 406
DB 1810 CCCAGCTCAGAGCCAGGTCCAAAGATTAACCAAGCCCTG-----CCCAAGAGAGAG 1757

QY 407 -----ProSerThrHisProLysGluLysAlaPheSerSerGluLysAspL 423
DB 1756 AAGTATTGCTTCTCTTAAGAGCAGCAGAGAGAG-----CAGAGGTGAG----- 1710

QY 423 euProTyrLeuSerThrThrGluMetTyrLeuCysArgTyr-His-----Gln 438
DB 1709 -----TGGCAGCCCAACATTAATAAC 1688

QY 439 ProProProSerProLeuProLeuArgLysSerSerProLysLysGluLysLysValAla 458
DB 1687 CACCTGAGAGTCCAGAGAGCCCTCAGAGTGTCTCATTAAGAGAGAGAGAGAGAGAGAGAG 1628

QY 459 ArgCysLeuMetProSerSerValAlaGlyGlnThrSerValLeuAlaValProSer--- 477
DB 1627 TCCACTTATTGGGGTCCCTAGGTGGCCCGTCTATCTGTAACTTGACTCTGTCTCTC 1568

QY 478 -----TyrArgAsp-----HisSerValGlu 484
DB 1567 TACATGTCTTTAATAGGGTGAAGGTGGAGAGAGAGAGAAATTATTTTCAGTGGCAAC 1508

QY 485 ProLeuArgAspProAsnProSer-----AspLeuLeuGluLysLeu 498
DB 1507 CCC---AATGACCCCAATCCAACTCAGGTGAGAGATGGCTCTTACTTGAAGAGACTA 1454

RESULT 13

US-09-081-385-6
; Sequence 6, Application US/09081385
; Patent No. 6593456

GENERAL INFORMATION:
 APPLICANT: Gatanaga, T.
 APPLICANT: Gatanaga, T.
 TITLE OF INVENTION: Factors Altering Tumor Necrosis
 TITLE OF INVENTION: Factors Altering Tumor Necrosis
 TITLE OF INVENTION: Factor Receptor Releasing Enzyme Activity, and Methods
 NUMBER OF INVENTION: of Use Thereof
 NUMBER OF SEQUENCES: 154
 CORRESPONDENCE ADDRESS:
 ADDRESS: MORRISON & FOERSTER
 STREET: 755 PAGE MILL ROAD
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304-1018
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows
 SOFTWARE: FASTSEQ for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/081,385
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/964,747
 FILING DATE: 05-NOV-1997
 APPLICATION NUMBER: 60/030,761
 FILING DATE: 06-NOV-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Wu, Frank
 REGISTRATION NUMBER: 41,386
 REFERENCE/DOCKET NUMBER: 22000-20577.21
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-813-5600
 TELEFAX: 650-494-0792
 TELETYPE: 706141
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3117 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 US-09-081-385-6

Alignment Scores:
 Pred. No.: 0.00203 Length: 3117
 Score: 223.50 Matches: 185
 Percent Similarity: 31.67% Conservative: 76
 Best Local Similarity: 22.45% Mismatches: 210
 Query Match: 6.90% Indels: 353
 DB: 4 Gaps: 42

US-10-054-935-2 (1-614) x US-09-081-385-6 (1-3117)

QY 12 AAlaAlaProAlaGlyGly-----AanProGluGluArgLeuAspArgGluAlaAla--- 28
 DB 293 GCTGTCCACAGAGTGGCTCTCAGTAAAGTGGCCCACTCTGTGTGCTCCAGGGCGGG 352
 QY 29 -----AlaAlaLeuGlyGlyPro 34
 DB 353 CCCGAGCGTGGAGAGGTGGGGGTGTCACTGACAGCAGCGTGGCAGCAGCAGCGCCCA 412
 QY 35 GluAspGluProGlyAlaAlaGluAlaHisPheLeuProArgHisArgGlyLeuLys--- 53
 DB 413 GCCTCCACCCCATTCATCAATGAA-----CTGCCACAGCTCTGTCCCTTACAGTGC 463
 QY 54 -----GluProGly-----ProProLeuAla 60
 DB 464 AACCAAGAGGAGCCCGATCTCTGAGTGGAGTCCCATCTTACTATACCACTCTGAGGC 523
 QY 61 SerSerGlnGly-----GlySerProAlaProSerPro----- 71

DB 524 ACTGAACGGGAGAAAGCGGGGGGCCACAGCTGACCCGCTATGTGCAACATGATGCC 583
 QY 72 -----AlaGlyCysGlyGlyLysGlyArgGlyLeuLeuProAlaGlyAlaAlaPro 89
 DB 584 ACAGAGGTGCACTGAGGTAGG-----GGGGCCCA 616
 QY 90 GlyGlnGlnGlu---GluSerTrpGlySerValProLeuProCysProProProAla 108
 DB 617 GGCACCCCTGAATTCCTTCCACGACGCAAGAAACCCCAACACAGTCACTGCCCTGCA 676
 QY 109 ThrLysGlnAlaGlyLe-----GlyGlyGluProAlaAlaAlaGlyAlaGly---Cys 125
 DB 677 ACCCTTCAGCTGGCATTGGCCACCAAGGAAACCGAGCTTCCGGAGGCCACC 736
 QY 126 SerProArgPro-----LysTrpGlnAlaValLeuProLleGlnThr 139
 DB 737 GCCCCCAACCCGGTGGCTGCTTCCCTCCAGAAAGCAGCAGCAGCAGCAGCAGCAGC 796
 QY 140 GlySerLeuValAlaAlaAlaLysGluProThrPro----- 151
 DB 797 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 856
 QY 152 -----TrpAla 153
 DB 857 TTCATGACCAAGCAACCTCGCAGCAACCCAGACTTTGGCTGACGACGCTGGCC 916
 QY 153 ----- 153
 DB 917 ACTGGACAGTCCACCTGCTGCTACCAACAGATGGACCTTACCCTTCCCCCAACC 976
 QY 154 -----GlyAspLysGlyGlyAlaAlaSerProAlaAlaThr 165
 DB 977 AGATATGAACCCAGAACTGGCAAGCCCTTTCAGAGA---CTCAGCCCGCAGCAGC 1033
 QY 166 AlaSerAspProAlaGlyProProProLeuProGlyProProProLeu----- 183
 DB 1034 GCTACC-----TCAGTCCAGATCCCTTCCCGCCGCTCCGCTCTTAAGA 1087
 QY 183 ----- 183
 DB 1088 GGGTATCTGCTCCACGAGCCCTGGATGGGGCTGGACCCAGCTGGGAGAGGCCAC 1147
 QY 183 ----- 183
 DB 1148 TGGCAACCTGTTCTATCATCATCGGCCCTGACAGACCGCCCACTGCTCCCTGGGGCA 1207
 QY 184 AlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlyGly----- 197
 DB 1208 GGGCATCTCTGAAGCTTGGGATTCCTGCTGAGAGTGGAGTCCAGCTACTGCTGTA 1267
 QY 198 -----ArgTrpLysSerMetArgLysSerProLeuGlyGlyGly 211
 DB 1268 TGGGAGAGACTAGCAACCAATGCGCGGAGCGAGAGGCTCTGCCATGGGAGCGAGGA 1327
 QY 212 GlySerGlyAlaSer-SerGlnAlaAlaCysLeuLysGlnLleLeuLeuLeuLeu 231
 DB 1328 GGGCATGAGGAGCTAGCAGCAGGAGCTGTGGCAG-----CTGCTTACGGGGCG 1378
 QY 231 pleuileGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 251
 DB 1379 AGTATCCAGAGCAGCAGCAGGAGCGCGGCA----- 1412
 QY 251 sSerGluArgAspThrLeuLeuAlaArgLleGluArgMetGluArgMetGlnLeuVal 271
 DB 1413 -TCCAGAGGCCAATTTGCTGACC-----CTGGCCCAAGAGGCTGTGAGCTGCG 1462
 QY 271 LysLysAspAspGluLysGluArgHisLysLysLeuPheGlnGlyTrpGluThrGluGlu 291
 DB 1463 CTCACGCAAGAAATGCAAGAT-----GGCAGGTCTTGAAGAGAA 1504
 QY 291 gGluGluThrGluLeuSerGlyLysLysLysLysLysGluCysGlnProGluLeuSerGlu 311
 DB 1505 GCGAAGAGTATGCTTGGCTCACTACCAAG-----TGTGGGTGGAGTTTCTGAGCC 1558

QY 75 GlyGlyLysGlyArgGlyLeu-----Leu 82
Db 24267 GCCGCTCACTGTCGCGCGCATCGAGGACGCTCAAGCAGAGGTCCTCC 24208
QY 83 LeuProAlaGlyAlaAlaProGlyGlnGlnGlnSer-----TrpGly----- 97
Db 24207 CTTCCGCGCTGCGATCGACAGCGCGCGCGCGGAGCGGATGAGTACTGATGA 24148
QY 98 ---GlySerValProLeu-----ProCysProPro----- 106
Db 24147 AGAGGCGCAGCCGCTCGCGCGAGTGGCGGTGCACAGCTTGCCTCGCGCATGAGAGC 24088
QY 107 -----ProAlaThrLysGlnAlaGlyIleGlyGlyLysProAla----- 119
Db 24087 GCTGGAACACTGTCGCT-----GATCCGCGCGCGGAGTCGATGCCGT 24043
QY 120 AlaAlaGlyAlaGlyCysSerProArg-----ProLysTrpGlnAlaVal 134
Db 24042 GGTCTGAGACGTGATGCGCGCTCGCTCGGAGGCGCTCCACGTCACAGACGATGGGCT 23983
QY 135 LeuProIleGln-----ThryGlySerLeuValAlaAlaAlaLys 147
Db 23982 CTTCCGCGCGCGCTGTACTTGTGCGCTTCACAGAGTTTCACAGACCTGCTCCAGCC 23923
QY 148 GluProThrProTrpAlaGlyAspLysGlyGlyAlaAlaSerPro-----AlaAla 164
Db 23922 TGTCTCTGTCTCCGCGCGACCCACAGCGCTCACAGGCGACTTCACGCGCAGAGGCGCT 23863
QY 165 ThrAlaSerAspProAlaGlyProProProLeuProLeuProGlyProProProLeuAla 184
Db 23862 CGAAGCGCGCGGAGGTGTCCACCACTCCGCCACCAAGTTGCCACCTCCAGGCGG 23803
QY 185 ProThrAlaThrAlaGlyThrLeuAlaAlaSerGlyLysArgTrpLysSerMetAlaGlyLys 204
Db 23802 CCA-----GATCAGTGCAGCC----- 23785
QY 205 SerProLeuGlyGlyGlyGlyGlySerGlnAlaAlaCysLeuLysGln 224
Db 23784 TCCCGCGCTCCAGGCGCGGAC-GGCTCCAGCAGGTCATCCACAGCCCGCAGCCGATCC 23726
QY 225 IleLeuLeuLeuGlnLeuAspLeuIleGln----- 235
Db 23725 ACCTGCGCTTGACCTTGAGCACCGTGGC-CAAGTCCACCGCGTGGCGCGCAGCAGCGC 23667
QY 236 -----GlnGlnGlnGlnGlnAlaLysGlyLysGlyIleGlnGlnLeuLys 251
Db 23666 CGGCTTCATGTCGAGGCGCGCTTCAGCGCGCGTGAAGCGGTCTTCACTCGTGGCT 23607
QY 252 SerGluArgAspThrLeuLeuAlaArgIleGluArg----- 263
Db 23606 GCGGATGAGATGAATCTCTCGCGCACCTGAGCGCGCTTGAGCTCGCGCAGCAGCGC 23547
QY 264 -----MetGluArgArgMetGlnLeuValLysLysAsp-----Asn 275
Db 23546 CGCGTTCTCACAGCCACCGCGCACTGTTCGCGCGCGCTCCACAGTCCAGCTCCG 23487
QY 276 GluLysGlnArgHisLysLeuPhe-----GlnGlyTrpGlnThrGlnGln 290
Db 23486 CACAGAGAAGAGAGTGCCTCTTGTGAGACAGACAGCAGCCACCTCGCGCGCG 22427
QY 291 ArgGlnGlnThrGlnLeuSerGlyLys-----IleLysLeuGln 303
Db 23426 CGCGCACAGCGGACAGCGCGCAAAATGAGCCCATGAAATCCCATAGCCCGCTGAGT 23367
QY 304 CysGlnProGlnLeuSerGlnThrSerGlnThrLeuProProLysProPheSerCysGly 323
Db 23366 GCCCACTGCGCTG-----ACGCA-GCGCAGCGCTGCGCGCAGAGCGGT 23323
QY 324 ArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArgLysThrPro 343
Db 23322 CCGACTGGG-----AGTCGAGAGTGTGCTCGGCGGACTCCGAGCGCGCAGGTCG 23272
QY 344 ValLysLysLeuAlaProGlnPheSerLysValLysThrThrProLysHisSerPro 363

Db 23271 -----ACACCG----- 23266
QY 364 IleLysGlnGlnProCysGlySerLeuSerGlnThrValCysLysArgGlu----- 380
Db 23265 -----CACCGCGCGCAGGCTCCCGCTCTCTTCGCGAGAGAAAGACTCGG 23218
QY 381 -----LeuArgSerGlnGlnThrProGluLysPro-----ArgSerSerValAspThrPro 397
Db 23217 CGCGCGCACGTGCGCAGCGGACAGGCGCACACAGCGGCGCGCGCAGACGTGTCCA 23158
QY 398 ProArgLysSerThrProGlnLysGly-----ProSer 408
Db 23157 CTTCAAGCTCTTCCCGACCGGCAATCTCTTCGACCGCTCGGAGAGGCCCT 23098
QY 409 ThrHisProLysGlnLysAlaPheSerSerGlnIleGluAspLeuProTrpLysSerThr 428
Db 23097 CGTGTCACAGCGCGCAGGCT----- 23077
QY 429 ThrGluMetTrpLeuCysArgTrpHisGlnProProProSer-----ProLeu 444
Db 23076 -----CCACACCAACAGACCGCGCTGCTCGC 23047
QY 445 ProLeuArgGluSerSerProLysLysGlnGlnThrValAlaArgCysLeuMetProSer 464
Db 23046 CGCGCAGCGCGGCTTCACTTCACCGGACCGGACCTGACGACGCGCGCG-CTGATGACGTGG 22988
QY 465 SerValAlaGlyGlnThrSerValLeuAla-ValProSerTrpArgAspHisSerValGln 484
Db 22987 CCGGTATGGAGACGACCGTCCCGATGCGCGCTCCAGAGGCGTCCAGACTCCGCGTGGC 22928
QY 484 uProLeuArgAsp-----ProAsnProSerAspLeuLeuGln 496
Db 22927 TCCAGGCTCTCCCGCATGTGCGCACCGCGGTGACGCGCCAGCGCTCT----- 22879
QY 496 uAsnLeu-AspAspSerValPheSerLysArgHisAlaLysLeuGlnLeuAspGluLys 516
Db 22878 ----TCCGCGCGAATCTTCTTCGACACCTGACCGGACGAGTCCGAGACCGCAAC 22823
QY 516 rGArgLysArg-----TrpAspIleGlnArgIleArgGlnGlnArgI 530
Db 22822 AGCGCAGCGGACGAGCAGATATGATGATCTTCGACGTCGAGGCGGAGG----- 22779
QY 530 IleLeuGlnArgLeuGlnLeuArgMetTrpLysLysLysGlyIleGlnGlnSerGlnProG 550
Db 22778 -----AAACCGCATGTGCTCGCGCAGCGGAGGAGAGG 22748
QY 550 IuValThrSerPhePheProGluProAspAspValGluSerLeuMetIleThrProPheL 570
Db 22747 CTCCCTTGAATTATGGGCTCGCGCGAG----- 22719
QY 570 euProValAlaAlaPheGlyArgProLeuProLysLeuThrProGlnAsnPheGlnLeuP 590
Db 22718 -----CGGCAACCGGCGCC-----CCGAG-----GATTCG 22691
QY 590 rOTrPLeuAspGluArgSerArgCysArgLeuGlnIleGlnLysLysGlnThrProHisA 610
Db 22690 CCTGTGTGAGCC-CGGCGCGTGTGGCGGCGGAGATCGAAGAGGATGAACCGCGG 22632
QY 610 rGThrCys 612
Db 22631 GTTGTGT 22624

Search completed: March 25, 2005, 23:08:27
Job time : 375 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 25, 2005, 21:07:33 ; Search time 890 Seconds
(without alignments)
4083.956 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 3238
Sequence: 1 MTRSAVFRAAAGAGNPE.....RSRCRLRIQKQPHPTCRK 614

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus_p2n.model -DEY=xlh
-Q=cgnt2_1/USPTO.spool/US10054935/rnatc_23032005_144819_26082/app_query.fasta_1.775
-DB=N.Geneseq_16Dec04 -QPM=faetap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=5 -DOCLIGN=200 -THR SCORE=pcpt -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US10054935 @CGN 1.1 470 @rnatc_23032005_144819_26082 -NCPH=6 -ICPH=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq_16Dec04:.*
1: geneseqn19808:.*
2: geneseqn19908:.*
3: geneseqn20008:.*
4: geneseqn20018:.*
5: geneseqn20028:.*
6: geneseqn20038:.*
7: geneseqn20048:.*
8: geneseqn20058:.*
9: geneseqn20068:.*
10: geneseqn20078:.*
11: geneseqn20088:.*
12: geneseqn20098:.*
13: geneseqn20108:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3238	100.0	4372	8	ACC72044 BCUI041 g
2	2406.5	74.3	1755	12	Adq86569 Human tum
3	2406.5	74.3	1755	13	Adq84001 Human tum
4	2406.5	74.3	1755	13	Adq83399 Human tum
5	1967.5	60.8	2106	10	Adf59738 Human con

6	1448.5	44.7	1529	3	AAf18193	AAf18193 Lung canc
7	1408.5	43.5	1143	10	Adf57688	Adf57688 Human pol
8	1351.5	41.7	2887	3	AAc76874	AAc76874 Human ORF
9	1055.5	32.6	615	6	ABK62818	ABK62818 Rat seque
10	1053.5	32.5	1290	11	AD130986	AD130986 Human cDN
11	819	25.3	3425	12	AD064204	AD064204 Novel hum
12	680.5	21.0	2887	3	AAc76874	AAc76874 Human ORF
13	625	19.3	365	9	ACH42852	ACH42852 Human foe
14	269.5	8.3	865	12	ADP84445	ADP84445 Human bre
15	254.5	7.9	5973	10	ADCP9116	ADCP9116 Human KPP
16	249	7.7	3750	13	AD086035	AD086035 Human tum
17	246	7.6	3774	10	ADK11436	ADK11436 Human neu
18	246	7.6	3774	12	ADQ19719	ADQ19719 Human sof
19	246	7.6	7713	4	AAK51958	AAK51958 Human pol
20	245	7.6	5373	6	AAQ30567	AAQ30567 Human kin
21	245	7.6	5373	12	AD015093	AD015093 Human can
22	245	7.6	5438	12	ADJ96545	ADJ96545 Human dys
23	242	7.5	5619	12	AD140903	AD140903 Human kin
24	241	7.4	6242	10	AD071201	AD071201 Human int
25	240	7.4	7033	12	AD019537	AD019537 Human PRO
26	239.5	7.4	4755	13	ACN37449	ACN37449 Tumour-as
27	238	7.4	2325	12	AD023872	AD023872 Human sof
28	238	7.4	9551	2	AAZ22301	AAZ22301 cDNA enco
29	236	7.3	7033	3	AAZ36989	AAZ36989 Human per
30	235.5	7.3	4037	4	AAK52942	AAK52942 Human pol
31	235	7.3	4707	12	ADP95101	ADP95101 Human ser
32	233.5	7.2	2800	12	ADK52037	ADK52037 Human ato
33	233.5	7.2	2800	13	AD080296	AD080296 KIAA1543
34	233.5	7.2	7078	12	AD059147	AD059147 Angiococ
35	232.5	7.2	5858	8	AAQ55726	AAQ55726 Nephtla m
36	232	7.2	4664	8	ABX12008	ABX12008 Transport
37	231	7.1	3608	6	AAH33243	AAH33243 Human sec
38	230.5	7.1	7065	9	ADB78878	ADB78878 L. major
39	230.5	7.1	7065	9	ADB78878	ADB78878 Leishmani
40	230	7.1	22970	10	ADK70082	ADK70082 Mutant hu
41	230	7.1	22970	10	ADK70081	ADK70081 Wild type
42	229.5	7.1	2618	9	ACF35661	ACF35661 Human nuc
43	229.5	7.1	3591	13	ACN38689	ACN38689 Tumour-as
44	229.5	7.1	5328	12	AD085928	AD085928 Human tum
45	229	7.1	5373	12	ADP47970	ADP47970 Human PKI

RESULT 1

ACCT2044 standard; DNA; 4372 BP.

ALIGNMENTS

ID	Score	Query Match	Length	ID	Description
ACCT2044	3238	100.0	4372	8	ACC72044 BCUI041 g
ACCT2044	2406.5	74.3	1755	12	Adq86569 Human tum
ACCT2044	2406.5	74.3	1755	13	Adq84001 Human tum
ACCT2044	2406.5	74.3	1755	13	Adq83399 Human tum
ACCT2044	1967.5	60.8	2106	10	Adf59738 Human con

DR MPI: 2003-381623/36.
XX P-PSDB; ABR58310.
XX
XX New isolated human differentially-regulated breast cancer polynucleotide
PT and polypeptide, useful for diagnosing, staging, prognosticating,
PT preventing and/or treating diseases and conditions relating to breast
PT cancer.
XX
XX Claim 2; SEQ ID NO 65; 127bp + Sequence listing; English.
XX
XX The invention relates to isolated polynucleotides which are
CC differentially-regulated in breast cancer. The methods and compositions
CC of the present invention are useful for detecting, diagnosing, staging,
CC monitoring, prognosticating, preventing and/or treating diseases and
CC conditions relating to breast cancer, and may be used in gene therapy or
CC antisense therapy. They can also be used in research, drug discovery,
CC clinical medicine and forensic medicine. Sequences given in records
CC ACC72012-ACC72074 represent polynucleotides of the invention that are
CC differentially-regulated in breast cancer. NOTE: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 4372 BP; 1136 A; 1012 C; 1077 G; 1147 T; 0 U; 0 Other:

Alignment Scores:
Pred. No.: 1,84e-105 Length: 4372
Score: 3238.00 Matches: 614
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-054-935-2 (1-614) x ACC72044 (1-4372)
QY 1 MetThMetArgSerAlaValBhelYSAlaAlaAlaProAlaGlyGlyAenProGlu 20
DB 78 ATGACCATATGATCCCGGGGTTCAGAGCGCGCGCCCTCCGCGCAATCCCTGAG 137
QY 21 GlnArgLeuAspTyrGlnArgAlaAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
DB 138 CAGCGACTGAGCTACGAGCGGGCTGCGGCGCTGGGGGGCCGAGAGCAAGCTGGGGCG 197
QY 41 AlaGlnAlaHisPheLeuProArgHisArgGlyLeuGlyGluProGlyProProLeuAla 60
DB 198 GCCGGAAGCCCACTTCTCCCGCCGCAACCTCAAGAGCGCGGCGCCCGCTGCGCC 257
QY 61 SerSerGlnGlyGlySerProAlaProSerProAlaGlyGlyGlyGlyGlyGlyGly 80
DB 258 TCCCTCCAGAGCGCGAGCCCGCGCTTCCCGCGCGCTGCGCGCAAGGCGCGGCGC 317
QY 81 LeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGln 100
DB 318 TTGTTACTCCCG 377
QY 101 ProLeuProCysProProProAlaThrLysGlnAlaGlyTlleGlyGlyGluProAlaAla 120
DB 378 CCTTCCCTGTCGCA 437
QY 121 AlaGlyAlaGlyCysSerProAlaArgProLysTyrGlnAlaAlaLeuProLleGlnThrGly 140
DB 438 GCGGAGCG 497
QY 141 SerLeuValAlaAlaAlaLysGluProThrProTyrAlaGlyAspLysGlyGlyAlaAla 160
DB 498 TCTCTGTGGCG 557
QY 161 SerProAlaAlaThrAlaSerAspProAlaGlyProProProProLeuProProGlyPro 180
DB 558 TCCCG 617
QY 181 ProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlnGlyGlyGlyTyr 200

DB 618 CCAACCCCTGGCGCCACCGCCGAGACCTTGCGCGCCAGCGAGCGAGATGGAAG 677
QY 201 SerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAlaSerSerGlnAlaAla 220
DB 678 AGATATGAGAAAGACCCCTCTCGGGGGTGGCGGGCTCGGGAGCTCCAGTCAGCGCGCC 737
QY 221 CysLeuLysGlnLleLeuLeuLeuGlnLeuAlaPLeuLleGlnGlnGlnGlnGlnGln 240
DB 738 TGCCTCAACCAATCTTGTGCTGCAATGTGACCTCATGAAACAGAGAGCGAGCGCTG 797
QY 241 GlnAlaLysGlnLysGlnLleGlnGlnLysSerGlnArgAspThrLeuAlaAlaArg 260
DB 798 CAGCGCAAGAAAGAGATTCAGAGCTGAAGCTGAGAGACACGCTCTCTGCTCGG 857
QY 261 lleGlnArgMetGlnArgArgMetGlnLeuValLysLysAspAsnGlnLysGlnArgHis 280
DB 858 ATTGAACGTATGAAAGCGGATGACGTGTGTAAAGAAATGCCATTTGAAAGTAAAGAAAGGCAC 917
QY 281 LysLeuPheGlnGlyTyrGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
DB 918 AAGCTGTTTCAGGGCTATGAAACTGAAGAGAGAGAGAAACAGAGCTATCTGAGAAATTT 977
QY 301 LysLeuGlnCysGlnProGlnLeuSerGlnThrSerGlnThrLeuProProLysProPhe 320
DB 978 AAACCTGAGTGCAGCGCGAGCTTCCGAGACATCCCAAGCTTGCCTCCAGACCTTC 1037
QY 321 SerCysGlyArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGlnArg 340
DB 1038 TCATGTGGCGGAGGTGGAAGGGAATTAAGAAATCCCATTTGAAAGTAAAGAAAGA 1097
QY 341 LysThrProValLysLysLeuAlaProGluLysPheSerLysValLysThrLysThrProLys 360
DB 1098 AAGACTCTCTGTAAAAAGCTGCTCTGAAATTTCAAAAGTCAAAACAAAACTCTTAAG 1157
QY 361 HisSerProLleLysGlnGluProCysGlySerLeuSerGlnThrValCysLysArgGln 380
DB 1158 CACTCTCTTAAAGAGAAACCTGTGTTCTTATCTGAAACCTTTGTAAACGTGA 1217
QY 381 LeuArgSerGlnGlnThrProGlnLysProArgSerSerValAspThrProProArgLeu 400
DB 1218 TTGAGAGGCAAGAAACCCAGAAAGCCCGGTTCTTCAGTGAGACCCCAACAAAGCTC 1277
QY 401 SerThrProGlnLysGlyProSerThrHisProLysGlnLysAlaPheSerSerGlnLle 420
DB 1278 TCCACTCCCAAAAGGAGACCCAGACCACTCCCAAGAGAAAGACCTTCTCAAGTGAATA 1337
QY 421 GlnAspLeuProTyrLeuSerThrThrGlnMetTyrLeuCysArgTyrPheLysGlnProPro 440
DB 1338 GAGATTTGCCGTACTTTCACCAAGAAATATTTGTGTGTTGGCACAGCTCTCC 1397
QY 441 ProSerProLeuProLeuArgGlnSerSerProLysGlnGlnGlnThrValAlaArgCys 460
DB 1398 CATACACCGTTACATTAAGGAAATCTCTCCAAAGAAAGAGAGAGACTGAGCAAGGTGT 1457
QY 461 LeuMetProSerSerValAlaGlyGlnThrSerValLeuAlaValProSerThrPArgAsp 480
DB 1458 CTGATGCATCAAGTGTTCAGAGAGAACTTCACTGTGCTGTCTTCTTGAGAGGAC 1517
QY 481 HisSerValGlnProLeuArgAspProAsnProSerAspLeuLeuGlnLeuLeuAsp 500
DB 1518 CACTCACTGAGGCTCTTAAGGACCCCAATCTTCAAGCTTTTGAAGAACTTGAGATGAC 1577
QY 501 SerValPheSerLysArgHisAlaLysLeuGlnLeuAspGlnLysArgArgLysArgTyr 520
DB 1578 AGTGTGTTTTGGAAGGCGCATGCAAAACGTGAGCTGATGAGAAAGAAAGAAAGATGG 1637
QY 521 AspLleGlnArgLleArgGlnGlnArgLleLeuGlnArgLeuGlnLeuArgMetTyrLys 540
DB 1638 GATATTCAGAGGATCAGGGAACAAAGAAATTTTCAACGACTGCAAGCTCAGAAATGATAAA 1697
QY 541 LysLysGlyTlleGlnLysSerGlnProGlnValThrSerPhePheProGlnProAspAsp 560
DB 1698 AAGAAAGAAATTCAGAAATCTGAGCTGAGGTAACTTATTTTCCCTGAGCGAGATGAT 1757

QY 561 ValGluSerLeuMetIleThrProPheLeuProValAlaPheGlyArgProLeuPro 580
DB 1758 GTTGAAGTTGATGATTACCCCTCTCTGTTGTAGCATTTGGACACCATTTCCA 1817
QY 581 LysLeuThrProGlnMetPheGluLeuProTrpLeuAspGluArgSerArgCysArgLeu 600
DB 1818 AATTTAATCTCCAGAAATTTTGAGTACCTGGTGGATAGCGGTAGCCGATGCAGATTG 1877
QY 601 GlnIleGlnLysLysGlnThrProHisArgThrCysArgLys 614
DB 1878 GAGATCCAGAAAGCAACACCTCACCGACGTTGAGAAA 1919
RESULT 2
ADQ86569
ID ADQ86569 standard; cDNA; 1755 BP.
XX
AC ADQ86569;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human tumour-associated antigenic target (TAT) cDNA sequence #3442.
XX
DE human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
XX cancer; cell proliferative disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN MO2004060270-A2.
XX
PD 22-JUL-2004.
XX
PF 15-OCT-2003; 2003MO-US029126.
XX
PR 18-OCT-2002; 2002US-0418988P.
XX
PA (GETH) GENENTECH INC.
PA (MUTD/) MU T D.
PA (ZHOU/) ZHOU Y.
XX
PI Wu TD, Zhou Y.
XX
DR WPI; 2004-534300/51.
XX
PT New nucleic acid molecule and encoded polypeptide, for diagnosing,
XX preventing or treating cell proliferative disorders such as cancer.
XX
PS Claim 1; SEQ ID NO 3442; 5504bp; English.
XX
CC The present invention describes an isolated tumour-associated antigenic
CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
CC sequence identity to (a) - (c); or (e) a sequence that hybridises to (a) -
CC (c). Also described: (1) an expression vector comprising the above
CC nucleic acid; (2) a host cell comprising the above expression vector; (3)
CC a process for producing a polypeptide; (4) an isolated polypeptide
CC comprising: (a) an amino acid sequence encoded by any of the above
CC nucleotide sequences; (b) an amino acid sequence encoded by the full-
CC length coding region of the above nucleotide sequences; or (c) a sequence
CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
CC comprising the above polypeptide fused to a heterologous polypeptide; (6)
CC an isolated antibody that binds to the above polypeptide; (7) a process
CC for producing the antibody; (8) an isolated oligopeptide that binds to
CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)
CC binding organic molecule that binds to the above polypeptide; (10) a
CC composition of matter comprising the above (chimeric) polypeptide,
CC antibody, oligopeptide or TAT binding organic molecule, in combination
CC with a carrier; (11) an article of manufacture comprising a container and
CC the composition of matter contained within the container; (12) methods of
CC inhibiting the growth of a cell that expresses the above protein, where
CC the growth of the cell is at least in part dependent upon a growth
CC potentiating effect of the above protein; (13) a method of

CC therapeutically treating a mammal having a cancerous tumour comprising
CC cells that express the above protein; (14) a method of determining the
CC presence of a protein in a sample suspected of containing the protein
CC described above; (15) methods of diagnosing the presence of a tumour in a
CC mammal; (16) a method for treating or preventing a cell proliferative
CC disorder associated with increased expression or activity of the above
CC protein; and (17) a method of binding an antibody, oligopeptide or
CC organic molecule to a cell that expresses the protein described above.
CC The TAT sequences have cytostatic activities, and can be used in gene
CC therapy. The composition and methods are useful for diagnosing,
CC preventing or treating cancer. The composition is also used for preparing
CC a medicament for the therapeutic treatment or diagnostic detection of a
CC cell proliferative disorder or cancer. The present sequence represents a
CC human TAT cDNA sequence from the present invention.
XX
SQ Sequence 1755 BP; 435 A; 502 C; 505 G; 313 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 1,48e-76 Length: 1755
Score: 2406.50 Matches: 462
Percent Similarity: 97.27% Conservative: 1
Best Local Similarity: 97.06% Mismatches: 6
Query Match: 74.32% Indels: 7
DB: 12 Gaps: 2
US-10-054-935-2 (1-614) x ADQ86569 (1-1755)
QY 3 MetArgSerAlaValPheLysAlaAlaAlaProAlaGlyArgProGlnArg 22
DB 1 ATGAGATCCCGGGTGTTCAGAGCGCGCCGCTGCGCGCAATCTGAGCACCGA 60
QY 23 LeuAspTrpGluArgAlaAlaAlaLeuGlyGlyProGluAspGluProGlyAlaAlaGlu 42
DB 61 CTGGACTACAGAGCGGCTGGCGGCTGGGCGGCGCGAGACAGCTGGGGCGCGGAA 120
QY 43 AlaIlePheLeuProArgHisArgLysLeuLysGluProGlyProProLeuAlaSerSer 62
DB 121 GCCCACTTCTCCCGCGCACCGTAACTCAAGAGCGGGGCGCCGCTGCTCTCC 180
QY 63 GlnGlyGlySerProAlaProSerProAlaGlyCysGlyGlyLysGlyAlaLeuLeu 82
DB 181 CAGGCGGAGGCGCCCGCTTCCCGCGCGCTGCGCGCAAGGCGCGCTTGTTA 240
QY 83 LeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 102
DB 241 CTCCCGCTTG 300
QY 103 ProCysProProProAlaThrLysGlnAlaGlyLysGlyGlyGlyGlyGlyGly 122
DB 301 CCTGTCCCG 360
QY 123 AlaGlyCysSerProArgProLysGlyGlyGlnAlaValLeuProIleGlnThrLysSerLeu 142
DB 361 GCGCGCTGCTTC 420
QY 143 ValAlaAlaAlaLysGluProThrProTrpAlaGlyAspLysGlyGlyAlaAlaSerPro 162
DB 421 GTGGCTTCCC 480
QY 163 AlaAlaThrAlaSerAspProAlaGlyProProProProLeuProGlyProProPro 182
DB 481 GTCGCCACCGCTCGGACCGCGGCGGACCGCGCACCACTACCTGCGCGGCGCGCACCC 540
QY 183 LeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGlyAlaGlyTrpLysSerMet 202
DB 541 CTGGCATGAAAGATAG 600
QY 203 ArgLysSerProLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 222
DB 601 AGGAAGAGCCCTCTCGGGGCTGTGGCGGCTCGGAGGCTTCAGTCAAGCGCGCTGCTC 660
QY 223 LysGlnIleLeuLeuLeuGlnLeuAspLeuIleGlnGlnGlnGlnGlnGlnGlnGln 242

```

Db 661 AACAGATCCTTCTGTCGATTGGAACCTCATCGAAGACAGACAGCAGCTGCAGGCC 720
Qy 243 LysGluLysGluLysGluLysSerGluArgAspThrLeuAlaArgLysGlu 262
Db 721 AAGGAAAGAGATGAGAGAGCTGAAGTGAAGAGAGACACGCTCTGCTCGGATTTGAA 780
Qy 263 ArgMetGluArgArgMetGlnLeuValLysLysAspArgGluLysGluArgLysLeu 282
Db 781 CGATAGAAAGCGGATGCGCTGCTGAAGAAAGATGAAGAAAGAAAGGCAAGCTG 840
Qy 283 PheGlnGlyThrGluThrGluLysGluLysGluLysGluLysGluLysLeu 302
Db 841 TTTTCAGGGGTATGAAGACTGAAGAGAGAGAGAAACAGACTATCTAGAAATTAACG 900
Qy 303 GluGluGlnProGluLeuSerGluThrSerGlnThrLeuProPheLysProPheSer 322
Db 901 GAGTCCAGCCGAGAGCTTCCAGAGACATCCAGACTCTGCTCCCAAGCCCTTCATG 960
Qy 323 GlyArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArgLysThr 342
Db 961 GGGCGGAGTGAAGAGGACATAAAGAAATCCCACTTTGGAAGTACAGAAAGAAAGACT 1020
Qy 343 ProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysHisSer 362
Db 1021 CCTGTTAAAGACTGGCTCTGTAATTTTCAAAAGTCAAAACAAACTCTTAAGCACTCT 1080
Qy 363 ProLysGluGluProGluProGluSerLysLeuSerGluThrValCysLysArgLysLeuArg 382
Db 1081 CCTATTAAAGAGAAACCTGTGGTCTCTTATCTGAAGACTGTTGTAACGTGAATTAAGAG 1140
Qy 383 SerGlnGluThrProGluLysProArgSerSerValAspThrProProArgLeuSerThr 402
Db 1141 ACCCAAGAAACCCCAAGAAAGCCCGGCTTCAAGTGAACAACCCCAAGAACTCTCACT 1200
Qy 403 ProGlnLysGluProSerThrHisProLysGluLysAlaPheSerSerGluLysLeuArg 422
Db 1201 CCCCAAGAGGACCCAGACACCATCCCAAGAGAAAGCCCTTCAAGTGAAGTGAAGAT 1260
Qy 423 LeuProLysLeuSerThrThrGluMetLysLeuGlyArgGlyThrPheGlnProProSer 442
Db 1261 TTGCGGTACCTTCCACACAGAAATGATTTGTGTCGTTGGACACAGCCTCCCATCA 1320
Qy 443 ProLeuProLeuArgLysSerSerProLysLysGluLysGluLysAlaArgCysLeuMet 462
Db 1321 CGGTAAACCTTACGGGAATCTCTCCAAAGAAAGAGAGAGACTGTACCA----- 1368
Qy 463 ProSerSerValAlaGlyLysThrSerValLeuAlaValProSerTrp 478
Db 1369 -----AGTAAGGCAATAGAAACACTTGCCTT---ATACCTTAGTGG 1407

```

```

PA (GENE) GENENTECH INC.
PA (WU/D/) WU T D.
PA (ZHOU/) ZHOU Y.
XX Wu TD, Zhou Y;
XX MPI, 2004-534300/51.
XX New nucleic acid molecule and encoded polypeptide, for diagnosing,
PT preventing or treating cell proliferative disorders such as cancer.
XX Claim 1; SEQ ID NO 815; 5504bp; English.
PS
XX
CC The present invention describes an isolated tumour-associated antigenic
CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
CC sequence identity to (a)-(c); or (e) a sequence that hybridizes to (a)-
CC (c). Also described: (1) an expression vector comprising the above
CC nucleic acid; (2) a host cell comprising the above expression vector; (3)
CC a process for producing a polypeptide; (4) an isolated polypeptide
CC comprising: (a) an amino acid sequence encoded by any of the above
CC nucleotide sequences; (b) an amino acid sequence encoded by the full-
CC length coding region of the above nucleotide sequences; or (c) a sequence
CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
CC comprising the above polypeptide fused to a heterologous polypeptide; (6)
CC an isolated antibody that binds to the above polypeptide; (7) a process
CC for producing the antibody; (8) an isolated oligopeptide that binds to
CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)
CC binding organic molecule that binds to the above polypeptide; (10) a
CC composition of matter comprising the above (chimeric) polypeptide,
CC antibody, oligopeptide or TAT binding organic molecule, in combination
CC with a carrier; (11) an article of manufacture comprising a container and
CC the composition of matter contained within the container; (12) methods of
CC inhibiting the growth of a cell that expresses the above protein, where
CC the growth of the cell is at least in part dependent upon a growth
CC potentiating effect of the above protein; (13) a method of
CC therapeutically treating a mammal having a cancerous tumour comprising
CC cells that express the above protein; (14) a method of determining the
CC presence of a protein in a sample suspected of containing the protein
CC described above; (15) methods of diagnosing the presence of a tumour in a
CC mammal; (16) a method for treating or preventing a cell proliferative
CC disorder associated with increased expression or activity of the above
CC protein; and (17) a method of binding an antibody, oligopeptide or
CC organic molecule to a cell that expresses the protein described above.
CC The TAT sequences have cytostatic activities, and can be used in gene
CC therapy. The composition and methods are useful for diagnosing,
CC preventing or treating cancer. The composition is also used for preparing
CC a medicament for the therapeutic treatment or diagnostic detection of a
CC cell proliferative disorder or cancer. The present sequence represents a
CC human TAT CDNA sequence from the present invention.
XX
SQ Sequence 1755 BP; 435 A; 502 C; 505 G; 313 T; 0 U; 0 Other;

```

```

Alignment Scores:
Pred. No.: 1,48e-76 Length: 1755
Score: 2406.50 Matches: 462
Percent Similarity: 97.27% Conservative: 1
Best Local Similarity: 97.06% Mismatches: 6
Query Match: 74.32% Indels: 7
DB: 13 Gaps: 2

```

US-10-054-935-2 (1-614) X AD084001 (1-1755)

```

Qy 3 MetArgSerAlaValPheLysAlaAlaAlaProAlaGlyGlyYanProGluGlnArg 22
Db 1 ATGAGATCCGCGGTGTTCAAGCGGCGCCCTGCGCGGCAATCTGAGCGCGCA 60
Qy 23 LeuAspTyrGluArgAlaAlaAlaLeuGlyGlyProGluLysProGluAlaAlaGlu 42
Db 61 CTGGACTACGAGGCGGCTCGGCGCTGCGGCGGCGGAGAGAGAGCTGCGGCGCGGAA 120
Qy 43 AlaHisPheLeuProArgHisArgLysLysLeuLysGluLysProGlyProProLeuAlaSerSer 62

```

```

Db      121  GCCACCTCTCTCCCGGCAACGTAAGAGCGGGGCGCCCGCTGCTCTCC 180
Qy      63  GlnGlySerProAlaProSerProAlaGlyCysGlyGlyLeuLeu 82
Db      181  CAGGGGAGAGCCCGGCTCTCCCGCGCGCGCGCGCGCGCGCTTTA 240
Qy      83  LeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGln 102
Db      241  CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 300
Qy      103  ProCysProProProAlaThrIleGlnAlaGlyIleGlyGlyGly 122
Db      301  CCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Qy      123  AlaGlyCysSerProAlaProIleGlnAlaValLeuProIleGln 142
Db      361  GCGGCGCTGACGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCT 420
Qy      143  ValAlaAlaAlaAlaGlyGlnProThrProAlaGlyAspGlyGly 162
Db      421  GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 480
Qy      163  AlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeu 182
Db      481  GCTGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Qy      183  LeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlnGly 202
Db      541  CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 600
Qy      203  ArgGlySerProLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 222
Db      601  AGGAAGAAGCCCTCTCGGCGCGCGCGCGCGCGCGCGCGCGCT 660
Qy      223  LysGlnIleLeuLeuLeuGlnLeuAspLeuIleGlnGlnGlnGln 242
Db      661  AAACAGATCTCTGCTGCAATTGACCTCATCGAACAAGACAGCAG 720
Qy      243  LysGlnLysGlnIleGlnGlnLysSerGlnLysGlnGlnGlnGln 262
Db      721  AAGGAAAGAGAGATCGAGAGCTGAAGTGAAGAGACACGCTCTG 780
Qy      263  ArgMetGlnArgArgMetGlnLeuValLysAspAsnGlnLysGln 282
Db      781  CGATGGAAGAGCGGATGCGCTGTAAGAGATTAAGAGAAAGAGCA 840
Qy      283  PheGlnGlyTyrGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGln 302
Db      841  TTTCAAGGCTATGAACCTGAAGAGAGAGAGAAACAGAGCTATCT 900
Qy      303  GlnCysGlnProGlnLeuSerGlnThrSerGlnThrLeuProPro 322
Db      901  GAATGCGAGCGGAGCTTCTCGAGACATCCCAAGCTGCTCCCAAG 960
Qy      323  GlyArgSerGlyLysGlyIleLysArgLysSerProPheGlySer 342
Db      961  GGGCGAGTGGAAAGGACATAAAGAGAAATCCCATTTGAGACTAC 1020
Qy      343  ProValLysLysLeuAlaProGlnPheSerLysValLysThrLys 362
Db      1021  CTGTATAAAGCTGGCTCTGATATTTTCAAAAGTCAAAACAAATCT 1080
Qy      363  ProIleLysGlnLysProCysGlySerLysSerGlnThrValCys 382
Db      1081  CTAATTAAAGAGAACCTGTGTCTTATCTGAATCTGTGTGAATTA 1140
Qy      383  SerGlnGlnThrProGlnLysProArgSerSerValAspThrPro 402
Db      1141  AGCCAGAAAGCCCGAGAAAGCCCGGCTTTCACTGAGACACCAAC 1200
Qy      403  ProGlnLysGlyProSerThrAspLysGlnLysGlnLysLysSer 422

```

```

Db      1201  CCCCCAAAGGAGACCCAGCACCATCCCAAGAGAAAGCTTCTCACT 1260
Qy      423  LeuProTyrLeuSerThrThrGlnMetTyrLeuCysArgTrpHisGln 442
Db      1261  TTGCGGACTTTCCACACAGAAATGTATTGTGTGTGGACAGCGCTCC 1320
Qy      443  ProLeuProLeuArgGlnSerSerProLysLysGlnGlnThrVal 462
Db      1321  CCGTTACATTACGGGAAATCTCTCCAAAGAGAGAGACTGTAGCA 1380
Qy      463  ProSerSerValAlaGlyGlnThrSerValLeuAlaValProSer 478
Db      1369  -----AGTAAAGCATATAGAGAACATTGCTTT---ATACCTTAGT 1407

RESULT 4
AD083399
ID  AD083399 standard; cDNA; 1755 BP.
AC  AD083399;
DT  07-OCT-2004 (first entry)
XX
DE  Human tumour-associated antigenic target (TAT) cDNA sequence #213.
XX
KW  human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
KW  cancer; cell proliferative disorder; gene; ss.
XX
OS  Homo sapiens.
XX
PN  WO2004060270-A2.
XX
PD  22-JUL-2004.
XX
PF  15-OCT-2003; 2003MO-US029126.
XX
PR  18-OCT-2002; 2002US-0418988P.
XX
PA  (GENTH ) GENENTECH INC.
XX
PA  (WUTD/) WU T D.
XX
PA  (ZHOU/) ZHOU Y.
XX
PI  Wu TD, Zhou Y;
XX
DR  WPI; 2004-534300/51.
XX
PT  New nucleic acid molecule and encoded polypeptide, for diagnosing,
PT  preventing or treating cell proliferative disorders such as cancer.
XX
XX  Claim 1; SEQ ID NO 213; 5504bp; English.
XX
CC  The present invention describes an isolated tumour-associated antigenic
CC  target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
CC  sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
CC  (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
CC  sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-
CC  (c). Also described: (1) an expression vector comprising the above
CC  nucleic acid; (2) a host cell comprising the above expression vector; (3)
CC  a process for producing a polypeptide; (4) an isolated polypeptide
CC  comprising: (a) an amino acid sequence encoded by any of the above
CC  nucleotide sequences; (b) an amino acid sequence encoded by the full-
CC  length coding region of the above nucleotide sequences; or (c) a sequence
CC  having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
CC  comprising the above polypeptide fused to a heterologous polypeptide; (6)
CC  an isolated antibody that binds to the above polypeptide; (7) a process
CC  for producing the antibody; (8) an isolated oligopeptide that binds to
CC  the above polypeptide; (9) a tumour-associated antigenic target (TAT)
CC  binding organic molecule that binds to the above polypeptide; (10) a
CC  composition of matter comprising the above (chimeric) polypeptide,
CC  antibody, oligopeptide or TAT binding organic molecule, in combination
CC  with a carrier; (11) an article of manufacture comprising a container and
CC  the composition of matter contained within the container; (12) methods of
CC  inhibiting the growth of a cell that expresses the above protein, where
CC  the growth of the cell is at least in part dependent upon a growth

```


CC potentiating effect of the above protein; (13) a method of
 CC therapeutically treating a mammal having a cancerous tumour comprising
 CC cells that express the above protein; (14) a method of determining the
 CC presence of a protein in a sample suspected of containing the protein
 CC described above; (15) methods of diagnosing the presence of a tumour in a
 CC mammal; (16) a method for treating or preventing a cell proliferative
 CC disorder associated with increased expression or activity of the above
 CC protein; and (17) a method of binding an antibody, oligopeptide or
 CC organic molecule to a cell that expresses the protein described above.
 CC The YAT sequences have cytostatic activities, and can be used in gene
 CC therapy. The composition and methods are useful for diagnosing,
 CC preventing or treating cancer. The composition is also used for preparing
 CC a medicament for the therapeutic treatment or diagnostic detection of a
 CC cell proliferative disorder or cancer. The present sequence represents a
 CC human YAT cDNA sequence from the present invention.

XX
 SQ Sequence 1755 BP; 435 A; 502 C; 505 G; 313 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,48e-76	Length:	1755
Score:	2406.50	Matches:	462
Percent Similarity:	97.27%	Conservative:	1
Best Local Similarity:	97.06%	Mismatches:	6
Query Match:	74.32%	Indels:	7
DB:	13	Gaps:	2

US-10-054-935-2 (1-614) x AD083399 (1-1755)

QY 3 MetArgSerAlaValPheIysAlaAlaAlaProAlaGlyIysAnPProGluInaArg 22
 Db 1 ATGAGATTCGGCGGTGTTCAAAGCGCGCGCGCCCTGCGCGGCAATCTTGACAGCA 60
 QY 23 LeuAspTyrGluArgAlaAlaAlaIleuGlyIysPProGluAspGluProGlyAlaAla 42
 Db 61 CTGACTACGAGGGGCTGGCGGCTGGCGGCGCGAGGACGAGCTGGGCGCGCA 120
 QY 43 AlaHisPheLeuProArgHisArgIysLeuIysGluProGlyProProleuAlaSerSer 62
 Db 121 GCCCACTTCCTCCCGCGACCGTAGCTCAAGAGACCGGGGCGCGCGCTCTCTCC 180
 QY 63 GlnGlyGlySerProAlaProSerProAlaGlyIysGlyIysGlyIysGlyIysLeu 82
 Db 181 CAGGGGGGAGCGCGCGCTTCCCGCGCGCTGGCGGCGCAAGGCGCGGCTTCTT 240
 QY 83 LeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 102
 Db 241 CTCGCCGCGGGGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 QY 103 ProCysProProProAlaThrIysGlnAlaGlyIleGlyIysGluProAlaAlaAla 122
 Db 301 CCTGTCTCCCG 360
 QY 123 AlaGlyCysSerProAlaProGlyIysGlnAlaValIleuProIleGlnThrGlySer 142
 Db 361 GCCGGGTGAGCT 420
 QY 143 ValAlaAlaAlaIysGluProThrProTyrAlaGlyAspIysGlyIysAlaAlaSer 162
 Db 421 GTGGCG 480
 QY 163 AlaAlaThrAlaSerAspProAlaGlyProProProProleuProGlyProProPro 182
 Db 481 GGTGCCACCG 540
 QY 183 LeuAlaProThrAlaThrAlaGlyIleThrIleuAlaAlaSerGlnGlyArgTyrPly 202
 Db 541 CTGGCG 600
 QY 203 ArgIysSerProleuGlyIysGlyIysGlyIysGlyIysGlyIysGlyIysGlyIys 222
 Db 601 AGGAAGAGCT 660
 QY 223 IysGlnIleLeuLeuLeuGlnIleuAspLeuIleGlnGlnGlnGlnGlnGlnGln 242

Db 661 AACCAATCTTGTGTCGAAATTGACCTTCAATGAAACAGACAGCAAGCTGCGAGCC 720
 QY 243 IysGluIysGluIleGluIleuIysSerGluArgAspThrIleuAlaArgIleGlu 262
 Db 721 AAGAAAGAGATGAGAGAGCTGAAGTCAAGAGAGAGACCGCTCTGCTCGATTGAA 780
 QY 263 ArgMetGluArgArgMetGlnIleuValIlyIysAspArgIlyIysGluArgHisIys 282
 Db 781 CGATGGAAGAGCGCGGTGACGTGTAAAGAGATACAGAAAGAAAGGCAAGCTG 840
 QY 283 PheGlnGlyTyrGluThrGluIleuArgGluGluIleuThrGluIleuSerGluIysIle 302
 Db 841 TTTCAGGCTATGAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 QY 303 GluCysGlnProGluIleuSerGluThrSerGlnThrIleuProProIysProPheSer 322
 Db 901 GAGTGCACCGGAGCTTCCGAGACATCCAGACTCTGCTCCAGAGCGCTTCTATG 960
 QY 323 GlyArgSerGlyIysGlyIlyIysArgIysSerProPheGlySerThrGluArgIysThr 342
 Db 961 GGGCGGAGTGAAGAGGACATAAAGGAATCCCATTTGAAATGACAGAAAGAAAGACT 1020
 QY 343 ProValIysIleuAlaProGluPheSerIysValIysThrIysThrProIysHisSer 362
 Db 1021 CCGTTAAAGAGCTGCTCTGATTTCAAAAGTCAAAACAAAACTCTTAAGCACTCT 1080
 QY 363 ProIleIysGluIleuProCysGlySerIleuSerGluThrValCysIysArgIleuArg 382
 Db 1081 CCAATTAAAGAGAAACCTGTGGTCTTATCGAAACGTTTGAAACGTGAATTGAAG 1140
 QY 383 SerGlnGluThrProGluIysProArgSerSerValAspThrProProArgIleuSerThr 402
 Db 1141 AGCCAAAGAAACCCAGAAAGAGCGCGCTTCAAGTGACACCCCAAGACTCTCACT 1200
 QY 403 ProGlnIysGlyIleuProIysHisIysPheIysGluIysAlaPheSerSerGluIleu 422
 Db 1201 CCCAAAG 1260
 QY 423 LeuProIleuSerThrThrGlnMetIysIysCysArgIysIysIysIysIysIys 442
 Db 1261 TTGCCGTACTTTCCACACAGAAATGTATTGTGTGCTGGACACAGCTCCCATCA 1320
 QY 443 ProLeuProleuArgIleuSerSerProIysIysGluIleuThrValAlaArgCysLeuMet 462
 Db 1321 CCGTTACCAATTCGGGAATCTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1368
 QY 463 ProSerSerValAlaGlyIleuThrSerValIleuAlaValProSerThr 478
 Db 1369 -----AGTAAAGCATAGAGAACTTGTCTT--ATACCTTAGTGG 1407

RESULT 5
 ID ADF59738 standard; cDNA; 2106 BP.
 XX ADF59738;
 AC 12-FEB-2004 (first entry)
 DT 12-FEB-2004 (first entry)
 XX
 DE Human contig polynucleotide sequence SEQ ID NO:2105.
 XX
 KW biological activity; genetic engineering; hybridisation probe; oligomer;
 KW primer; chromosome mapping; gene mapping; recombinant protein production;
 KW human; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2003080795-A2.
 XX
 PD 02-OCT-2003.
 XX
 PF 09-AUG-2002; 2002WO-US025485.
 XX

PR 09-AUG-2001; 2001US-0311261P.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Yang Y, Wang Z, Weng G, Ma Y;
 XX WPI; 2003-876918/81.
 DR P-PSDB; ADF60190.
 XX
 XX New polynucleotides, useful as hybridization probes, oligomers or
 PT primers, for chromosome or gene mapping, for the recombinant production
 PT of proteins, and for generating antisense DNA or RNA.
 PS Example 2; SEQ ID NO 2105; 571bp; English.
 XX
 CC The present invention describes isolated polynucleotide sequences (I),
 CC which encode polypeptides (II) with biological activity, also described:
 CC (1) a vector comprising (I); (2) an expression vector comprising (I); (3)
 CC a host cell genetically engineered to comprise (I) which is operatively
 CC associated with a regulatory sequence that modulates expression of (I) in
 CC the host cell; (4) a polypeptide (II) encoded by (I); (5) a composition
 CC comprising the polypeptide of (4) and a carrier; (6) an antibody directed
 CC against the polypeptide of (4); (7) detecting (I) or the polypeptide of
 CC (4) in a sample; (8) identifying a compound that binds to the polypeptide
 CC of (4); (9) producing the polypeptide of (4); and (10) a collection of
 CC polynucleotides comprising at least one of the polynucleotide sequences
 CC (I). The polynucleotides (I) can be used as hybridisation probes,
 CC oligomers or primers, for chromosome or gene mapping, for the recombinant
 CC production of proteins, and for generating antisense DNA or RNA. The
 CC present sequence represents a human contig polynucleotide sequence, which
 CC is used in an example from the present invention.
 XX
 SO Sequence 2106 BP; 526 A; 569 C; 588 G; 422 T; 0 U; 1 Other:
 Alignment Scores:
 Pred. No.: 4.6e-61 Length: 2106
 Score: 1967.50 Matches: 440
 Percent Similarity: 90.41% Conservative: 3
 Best Local Similarity: 89.80% Mismatches: 15
 Query Match: 60.76% Indels: 32
 DB: 10 Gaps: 4
 US-10-054-935-2 (1-614) x ADF59738 (1-2106)
 QY 10 AAlaAlaAlaProAlaGlyGlyAsnProGluGlnArgLeuAspTyrGluArgAlaAla 29
 DB 2 GCGGCGCGGCGCGCTGCGGCGCAATCTGAGCAGGACTGAGCTAGCGAGCGGCGGCG 61
 QY 30 AAlaGluGlyProGlu-AspGluProGlyAlaAlaGluAlaHisPheLeuProArgH 49
 DB 62 GCGCTGGGCGGCGCGCATGAGCAGAGCTGGGCGCGCGGAGCCCACTTCTCCCGGCA 121
 QY 49 SAArgLysLeuLysGlu-ProGlyProProLeuAlaSerSergLysGly-SerProAla 68
 DB 122 CCCTAAGCTCAAGAGCCGCGGCGCGCTGCGCTCTCCAGAGCGGAGAGCGCGCGC 181
 QY 69 ProSerProAlaGlyCys---GlyGlyLysGly-ArgLysLeuLeuLeuProAlaGly 87
 DB 182 CTTTCCCGGCGCGCGCGCGCGCGGCGGCGGCGGCGGCTGTCTCTCCCGCGCGGCG 241
 QY 87 AAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 107
 DB 242 GGGCGCGCGCGCGCGAG 301
 QY 107 AAlaThrLysGlnAlaGlyLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 127
 DB 302 GCGCACCACCAAGACCGCGGCAATGGGCGGAGCGCTGCGCGAGCGCGGAGCGCGG 361
 QY 127 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 147
 DB 362 CCGGCG 421
 QY 147 GGUProThrProTrpAlaGlyAspLysGlyGlyAlaAlaSerProAlaAlaThrAla 167

DB 422 AGAGCTTAGCGCTGTGCTGGGAGCAAGGCTGGGAGCTCCCGCGCTCCACCGCTC 481
 QY 167 rAaPProAlaGlyProProProLeuProLeuProGlyProProProLeuAlaProThrAl 187
 DB 482 GAGCGCGCGCGGAGCG 541
 QY 187 aThrAlaGlyThrLeuAlaAlaSerGluGlyArgTyrPheSerMetArgLysSerProle 207
 DB 542 CACCGAGCGGAGCG 601
 QY 207 uGlyGly-GlyGlyGlySerGlyAlaSerSergLysAlaAlaCysLeuLysGlnLeuLeu 227
 DB 602 CGGCGGCTGGGCGCGCGCTGGGAGCGCTCCAGTACGCGCGCGCTCTCAACAGATCTTC 661
 QY 227 euLeuGlnLeuAspLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 247
 DB 662 TGCTGCAATTGGAGCTCATGAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 721
 QY 247 leGluGlnLeuLysSergLysArgAspThrLeuLeuAlaArg-1leGluArgMetGluArg 266
 DB 722 TCAGAGAGCTGAAGTCAAGAGAGAGACCGCTCTGCTGTGANTGAACTGAGAAAGG 781
 QY 267 ArgMetGln-LeuValLysLysAspAspGlnLysGluArgHisLysLeuPheGlnGly 286
 DB 782 CGGATGAGGCTGTAAAGAGATTAACGAAAGAAAGCAGCAGCAGCTGTTTCAGGCGCTA 841
 QY 286 rGluThrGlnGluArgGlu-GluThrGlnLeuSerGluLys1leLysLeuGluCysGln- 305
 DB 842 TGAACCTGAAGGAG 901
 QY 306 ProGluLeuSer-GluThrSergLys-ThrLeuProProLysProPhe-SerCysGlyArg 324
 DB 902 CCGGAGCTTCCCGAGACATCCCAAGACTGCTCCCAAGCGCTTCTTATGTGGCGG 961
 QY 325 SerGly-LysGlyHisLysArgLysSerProPhe-GlySergThrGluArgLys---Thrp 343
 DB 962 AGTGAAAGAGGAGCATTAAGAGAAATCCCATTTGGGAGATACAGAAAGAAAGACTCC 1021
 QY 343 roValLysLysLeuAla-ProGluPheSerLysValLysThrLysThrProLysHisSer 362
 DB 1022 CTGTAAAGAGCTGGCTCTGTAAATTTCAAAAGTCAAAAGCAAAAGCTCTTAAGCACTCT 1081
 QY 363 ProIle--LysGluGlnProCysGlySer-LeuSer-GluThrValCysLysArgGluLe 381
 DB 1082 CCAATTAAAG 1141
 QY 381 uArg-SergLysLysThrProGluLysProArgSer-SerVal-AspThrProProArgle 400
 DB 1142 GAGGAGAGCAAGAAAG 1201
 QY 400 uSerThrProGlnLysGlyProSerThrHisProLysGluLysAlaPheSerSergLys 420
 DB 1202 CTCACCTCCCGCAAG 1261
 QY 420 eGluAspLeuProLysLeuSerThrThrGluMetCysLysCysArgTyrHisGlnPro-P 440
 DB 1262 AGAAGATTGCGCTTACCTTTCCACACAGAGAAAGTATTTGTGTGCTTGGGAGAGCTTC 1321
 QY 440 roProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrValAla---- 458
 DB 1322 CCCCATCAGCGTAAACATTAACGGAATCTCTCCAAAGAGAGAGAGAGAGAGAGAG 1381
 QY 459 -----ArgCys--LeuMetProSerSerValAla 467
 DB 1382 AGGCAATAGAAACACTGCGCTTAATACCTTAAGTGTGGCG 1425
 RESULT 6
 AAF18193 standard; DNA; 1529 BP.
 XX AAF18193;
 AC AAF18193;
 XX

DT 14-MAR-2001 (first entry)
XX Lung cancer associated polynucleotide sequence SEQ ID 212.
DE Human; lung cancer associated protein; neuroprotective; cytoskeletal;
XX cardioactive; immunomodulatory; muscular active; vulnary;
KW gastrocestinal; nephrotoxic; antiinfective; gynecological;
KM antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KM proliferative disorder; wound healing; infectious disease; ds.
OS Homo sapiens.
XX MO20005180-A2.
XX 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US005918.
XX 12-MAR-1999; 99US-0124270P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX Ruben SM,
XX WPI: 2000-587514/55.
XX P-PSDB; AAB58317.
XX Lung cancer associated gene sequences, referred to as lung cancer
PT antiGene, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer.
XX Claim 1; Page 676-677; 1425pp; English.
XX Polynucleotide sequences AAF1982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytoskeletal; cardioactive; and
CC immunomodulatory; muscular active general; vulnary; gastrocintestinal
CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the protein
CC or polynucleotide sequences. The lung cancer associated polynucleotide
CC sequences may be used for detection of lung cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The proteins may be used to treat disorders such as
CC neural, immune, muscular, reproductive, gastrocintestinal, pulmonary,
CC cardiovascular, renal, and proliferative disorders. The proteins may also
CC be used in the treatment of wounds and infectious diseases.
CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
CC used in the course of the invention for the identification and
CC characterisation of the polynucleotide and protein sequences
XX
SQ Sequence 1529 BP; 460 A; 344 C; 379 G; 344 T; 0 U; 2 Other;
Alignment Scores:
Pred. No.: 6.19e-43 Length: 1529
Score: 1448.50 Matches: 284
Percent Similarity: 95.64% Conservative: 1
Best Local Similarity: 95.30% Mismatches: 6
Query Match: 44.73% Indels: 7
DB: 3 Gaps: 2
US-10-054-935-2 (1-614) x AAF18193 (1-1529)
QY 181 ProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGlyA-GTTPlys 200
DB 3 CCAACCCCTGGCGCCACCGCCACCGCCGAGCCCTGGCGCCAGCAGGCAATGGAAG 62
QY 201 SerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAlaSerSerGlnAlaAla 220
DB 63 AGATAGAGGAAGAGCCCTCTCGGGGGTGTGTGGCGCTCGGAGACCTCCAGTCAAGCCGCC 122
QY 221 CysLeuLeuGlnLeuLeuLeuGlnLeuAlaPleuLeuLeuGlnGlnGlnGlnLeu 240

DB 123 TGCCTCAACAGATCTTCTGCTGCAATTGACCTCATGGAACAGACAGCAGCTG 182
QY 241 GlnAlaLysGluLysGluLeuGluGluLeuLysSerGluArgAspThrLeuLeuAlaArg 260
DB 183 CAGGCCAAGAAAGAGATTCAGGAGCTGGAAGTCAAGAGAGACACGCTCTTGCTCGG 242
QY 261 TleGluArgMetGluArgArgMetGlnLeuValLysLysAspAsnGluLysGluArgHis 280
DB 243 ATTGAACGTATGAAAGGCGGATGACGTGTAAAGAAAGATPACGAAAGAAAGGCAC 302
QY 281 LysLeuPheGlnGlyTyrGluThrGluGluArgGluGluThrGluLeuSerGluLysIle 300
DB 303 AAGCTCTTTCAGGCGCTATGAAACTGAGAGAGAGAGAAACAGAGTATCTAGAAAT 362
QY 301 LysLeuGluCysGlnProGluLeuSerGluThrSerGlnThrLeuProGluProPhe 320
DB 363 AAACGTAGTGCACGCGGAGCTTCCGAGACATCCAGACTCTGCTCCAAAGCCCTTC 422
QY 321 SerCysGlyArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArg 340
DB 423 TCATGTGGCGGAGTGGAAAGGACATMAAGAAATCCCATTTGGAAATACAGAAAGA 482
QY 341 LysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLys 360
DB 483 AAGACTCTCTTAAAAAGCTGCTCTGAAATTTCAAAAGTCAAAACAAAACTCTTAAG 542
QY 361 HisSerProLysGluGluProCysGlySerLysSerGluThrValCysLysArgGlu 380
DB 543 CACTCTCTATTAAAGAGAACCCGTGTGTTCTTATCTGAACCTGTTGAACCTGAA 602
QY 381 LeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProArgLeu 400
DB 603 TTGAGAGGCCAAGAAACCCAGAAAGCCCGTCTTCACTGACACCCACCAAGACTC 662
QY 401 SerThrProGluLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIle 420
DB 663 TCCACTCCCCAAGAGGACCCAGACCCATCCCAAGAGAAAGCCCTTCAAGTAGAAR 722
QY 421 GluAspLeuProTyrLeuSerThrThrGluMetTyrLeuCysArgGTPHisGlnProPro 440
DB 723 GAAGATTTCGCTTCCACACACAGAAATGATTTGTGTCGTTGGCAGCAGCTCCC 782
QY 441 ProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrValAlaArgCys 460
DB 783 CCATCAACCTTACATTAAGGAATCTCTCCAAAGAAAGAGAGACTGTAGCA----- 836
QY 461 LeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrp 478
DB 837 -----AGTAAAGCAATAGAGAACATTTGCTTT--ATACCTTAGTGG 875
RESULT 7
ID ADF57688 standard; cDNA; 1143 BP.
XX ADF57688;
AC ADF57688;
XX 12-FEB-2004 (first entry)
DT Human polynucleotide sequence SEQ ID NO:55.
XX biological activity; genetic engineering; hybridisation probe; oligomer;
KW primer; chromosome mapping; gene mapping; recombinant protein production;
KW human; gene; ss.
XX Homo sapiens.
XX OS
XX PN WO2003080795-A2.
XX 02-OCT-2003.
XX 09-AUG-2002; 2002WO-US025485.
XX

PR 05-APR-1999: 99US-0127728P.
 PR 30-MAR-2000: 200US-00540763.
 (CURA-) CURAGEN CORP.
 Shinkens RA, Leach M;
 PI Shinkens RA, Leach M;
 PI Shinkens RA, Leach M;
 DR MPI: 2000-602362/57.
 DR P-PSDB; AAB42665.
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 PS Claim 5; Page 4044-4045; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
 CC anticonvulsant; antidiabetic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antineoplastic; antibacterial;
 CC antiviral; antifungal; antineumatic; antihypertensive; antidiabetic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORF-associated disorder. The nucleic acids can be used to express ORF
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, anti-inflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 XX
 SQ Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 0 U; 2 Other:

Alignment Scores:
 Pred. No.: 2,65e-39 Length: 2887
 Score: 1351.50 Matches: 269
 Percent Similarity: 94.74% Conservative: 1
 Best Local Similarity: 94.39% Mismatches: 8
 Query Match: 41.74% Indels: 8
 DB: 3 Gaps: 2

US-10-054-935-2 (1-614) x AAC76874 (1-2887)

QY 194 AAsSerGluGlyArgTrpLysSerMetAlaGlySerProLeuGlyGlyGlyGlySer 213
 DB 5 GCCGGCGAGGCGAGATGAGAAAGATGAGAAAGAGCCCTTCGGGGGTGGTGGCGCTCG 64
 QY 214 GAlAsSerSerGlnAlaAlaCysLeuLysGlnIleLeuLeuLeuGlnLeuAAsPLeuIle 233
 DB 65 GGAAGC-TCAAGTACAGCCCGCTGCTCCCAACACATCTTGTGCTGCAATTGGACCTCATC 123
 QY 234 GAlngGlnGlnGlnGlnLeuGlnAlaAlaGlyGlyGlyGlyGlyGlyGlyGlyGly 253
 DB 124 GAACACAGACAGACAGAGCTGCAAGCCCAAGAAAGAGATGCAAGAGCTGAAAGTCAAG 183
 QY 254 ArgAspThrLeuLeuAlaArgIleGlyLysGlyMetGlyLysArgMetGlnLeuValLysLys 273
 DB 184 AGAGACACCTCTCTCTCGATTTGAACGATGAAAGGCGAGTGCAGCTGGTAAAGAG 243
 QY 274 AspAsnGlyLysGlyLysArgIleLysLeuPheGlnGlyTyrGlyLysArgGlyGlyGly 293
 DB 244 GATTAACGAGAAAGAGACCAAGCTGTTTCAAGGCTATATAAAGCTGAAGAGAGAGAA 303
 QY 294 ThrGluLeuSerGluLysIleLysLeuGlyCysGlnProGluLeuSerGlyThrSerGln 313
 DB 304 ACGAGCTATCTGAGAAATTAATTAACGTGAGTGCACGCGAGCTTTCCGAGACATCCAG 363

QY 314 ThrLeuProProLysProPheSerCysGlyArgSerGlyLysGlyLysLysArgLysSer 333
 DB 364 ACTGCTCTCCCAAGCCCTTCTCATGTGGCCGAGTGGAAAGGACATTAAGGAATCC 423
 QY 334 PropheGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSerLys 353
 DB 424 CCATTTGGAAAGTACGAAAGAAAGAAAGCTCTGTATTAAGAGCTGCTCTGTAATTTTCAAA 483
 QY 354 ValLysThrLysThrProLysHisSerProIleLysGlyGlnProCysGlySerLeuSer 373
 DB 484 GTCAAAACAAAACTCTTAAGCACTCTCTATTAAGAGAAACCTGTGTCTTATCT 543
 QY 374 GluThrValCysLysArgGlyLeuLeuArgSerGlnGlnThrProGluLysProArgSerSer 393
 DB 544 GAAACGTTGTAAAGTGAATTAAGAGACCAAGAAACCCAGAAAGCCCGGTCTTA 603
 QY 394 ValAspThrProProArgLeuSerThrProGlnLysGlyProSerThrHisProLysGln 413
 DB 604 GTGGACACCCCAAGCAAGCTCTCCACTCCCAAGAGGACCCAGACCCATCCCAAGAG 663
 QY 414 LysAlaPheSerSerGluIleGlyLysLeuProTyrLeuSerThrThrGluMetTyrLeu 433
 DB 664 AAGCCTTCTCAAGTGAATGAAGATTGCGGTACCTTTCCACACAGAAATGATTTG 723
 QY 434 CysArgTrpHisGlnProProProSerProLeuProLeuArgGlySerSerProLysLys 453
 DB 724 TGTGTTGACACAGCTCTCCCATCACCGTTACATTACGGGAATCTCTCCAAAGAG 783
 QY 454 GluGluThrValAlaArgCysLeuMetProSerSerValAlaGlyGluThrSerValLeu 473
 DB 784 GAGGAGACGTGAGCA-----AGTAAGGCAATAGAAACACTGTCTTT 825
 QY 474 AlaValProSerTrp 478
 DB 826 ---ATACCTAATGG 837
 RESULT 9
 ABK62818
 ID ABK62818 standard; cDNA, 615 BP.
 AC XX
 AC ABK62818;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 XX Rat sequence differentially expressed in response to a hepatotoxin #725.
 DE Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 OS Rattus norvegicus.
 XX
 PN WO200210453-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 30-JUL-2001; 2001WO-US023872.
 XX
 XX 31-JUL-2000; 2000US-0222040P.
 PR 02-NOV-2000; 2000US-0244880P.
 PR 11-MAY-2001; 2001US-0290029P.
 PR 15-MAY-2001; 2001US-0290645P.
 PR 22-MAY-2001; 2001US-0292336P.
 PR 06-JUN-2001; 2001US-0297457P.
 PR 13-JUN-2001; 2001US-0297457P.
 PR 19-JUN-2001; 2001US-0298884P.
 PR 09-JUL-2001; 2001US-0303459P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter MW, Johnson KR, Caestele AL, Elshoff MR;
 XX
 DR MPI, 2002-241625/29.

Db 660 GACAGTGTGTTTCGAGCGGCATGCAAACTGAGCTGATGAGAGAGAGAAAAGA 719
Qy 520 TPAPSLIEGLNARGILEARGIUGIINARQILEUGLNARGLEUGINLEUARGMETTYR 539
Db 720 TGGGATNTTCAGAGATTCAGGAGAACAAAGAAATTTTACAGGACTGCAGCTCAGAAATGTAT 779
Qy 540 LYSGLYSGLYIILEGINSUSERGIUPROGIUVALTHSERPHEPHEPROGIUPROASP 559
Db 780 AAAAAAGAAAGAAATTCAGGAATTCGAGCCTGAGCTTACCTATTTTCCCTGAGCCAGAT 839
Qy 560 AAGVALGUSERLEUWETIETHPROPHLEUPROVALAALAPHEGLYAGPROLEU 579
Db 840 GAGGTGAAAGTTGATGATTTACCCCTTCTTCCGTGTGTACATTGTGACACCATTTA 899
Qy 580 PROLYLEUTherProGlnAenPheGluLeuProTriPLeuAspGluArgSerArgCysArg 599
Db 900 CCAAAATTAATCTCAGAGAAATTTTGAGCTACCTGTTGATGATAGCTTACCATGACAG 959
Qy 600 LEUGLIILEGINSGLYSGINThrProHISARGThrCysArgLys 614
Db 960 TTGGAGATTCAGAGAAAGAAACACCTCAGCCGAGCGTATGAGAA 1004
RESULT 12
AAC76874/c
ID AAC76874 standard; cDNA; 2887 BP.
XX
AC AAC76874;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2429 polynucleotide sequence SEQ ID NO:4857.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antiparkeian; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antifungal; antibacterial; cancer; proliferative disorder; hyperextension;
KW antineuritic; gene therapy; cancer; proliferative disorder; hyperextension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; erythematoidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN MO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000MO-US008621.
XX
PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkete RA, Leach M;
XX
DR WPI; 2000-602362/57.
XX
PT P-PSDB; AAB42665.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 5; Page 4044-4045; 5507P; English.
XX

CC AAC74446 to AAC7606 encode, the proteins given in AAB40237 to AAB43387,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiparkeian; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
CC anticonvulsant; antirheumatic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antifungal; antipneumatic; antihypertensive; antineuritic; antineuritic;
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 0 U; 2 Other:
Alignment Scores:
Pred. No.: 1,02e-15 Length: 2887
Score: 680.50 Matches: 134
Percent Similarity: 87.74% Conservative: 2
Best Local Similarity: 86.45% Mismatches: 12
Query Match: 21,02% Gaps: 7
DB: 3 Gaps: 2
US-10-054-935-2 (1-614) x AAC76874 (1-2887)
Qy 324 ARGSEGLYLSGLYISLARGYLSERProPhGlySerThrGluArgLysThrPro 343
Db 2491 AGAGGGGGTGGAGGGGTTGACTTACCCCAATTTGGAGATGACAAAGAAAGACTCCT 2432
Qy 344 VALLYLEULENAPROGIUPHESERLYSVALYSEThrLysThrProLysHisSerPro 363
Db 2431 GTTAAAGAGTGGCTCTCTGATTTTCAAAAGTAAACAAAGAACTCTTACGACTCTCT 2372
Qy 364 ILEGLYGLNUPROCYSGLYSERLUSERGIUThrValCysLysArgGluLeuArgSer 383
Db 2371 ATTAAGAGGAACCTGTGGTCTTATCGAAAGCTTTGTAAAGTGAATGAGGAGC 2312
Qy 384 GINGLThrProGluLysProArgSerValAspThrProArgLeuSerThrPro 403
Db 2311 CAAAGAAACCCAGAAAGCCCGCTCTTCAAGTGAACCCCAAGAACTCTCACCTCC 2252
Qy 404 GINLYSGLYPProSerThrHisProLysGluValAlaPheSerSerGluIleuAspLeu 423
Db 2251 CAAAGAGAACCCAGAACCCATCCCAAGAGAAAGCTCTTCAAGTGAATGAGAAAGATTG 2192
Qy 424 PROTYLEUSERThrThrGluMetTYLEUCysArgTriHisGlnProProSerPro 443
Db 2191 CCTTACCTTTCACCAAGAAATTTGTGTGTGGACAGCTCCCATACCG 2132
Qy 444 LEUPROLEUARGIUSERSerProLysGluGluThrValAlaArgCysLeuMetPro 463
Db 2131 TTACCATTCAGGGAATCTCTCCAAAGAGAGAGACTGTACCA----- 2087
Qy 464 SERSERVALAGLYLThrSerValLeuAlaValProSerThrP 478
Db 2086 ---AGTAAGCATAGAGAACATTGCTT---ATACCTAGTGG 2048
RESULT 13
ACH42852
ID ACH42852 standard; cDNA; 365 BP.
XX
AC ACH42852;
XX
DT 13-OCT-2003 (first entry)
XX

Db 1906 TGCAGCTGCAGAGAGAAAAACGGAGGCTGAGCCGGAGCAGAGCGGCTAGAA----- 1959
QY 286 MetGlnLeuValIysLysAspIleuGluArgHisLysLeuPheGlnGly----- 285
Db 1960 GCAGAGCTGCCCGCAGGACGAGAGACAGACCGG-----CTGAGGGGTGAGCGG 2010
QY 286 TyrGluThrGluGluArgGluGluThrGluLeuSer----- 297
Db 2011 CGGGAGACGGAGACACTGGAGGCCGCTGCGACATCTCAGCTGGGTGAATGAT 2070
QY 298 GluLysIle---LysLeuGluCysGlnPro-----GluLeuSerGlu 310
Db 2071 GAGAAAGTCTCCAGAGCTACTGACAGCCTCGCACAAGATGGCAGAGAGCTGAGTCC 2130
QY 311 Thr-----SerGlnThrLeuProProLysProPheSerCysGluArgSerGly 326
Db 2131 TTGAGGACGTAGGACCCAGACGCTCCCTGCGCGGCACTGAAGATGGAGGCTCGGCC 2190
QY 327 LysGlyHisLysArg-----LysSerProPheGlySerThrGlu 339
Db 2191 AGGCTGAGCTGACAGCGCTGAGAGCGCAGATCCGGCAGAGAGGCTGCAGAG 2250
QY 340 ArgLysThrProValLysLys----- 346
Db 2251 CGGCTGACACAGGTGAGAGGCGCAGCTGACAGCGCGCTGCAGAGGCGCAG 2310
QY 347 -----LeuAlaProGluPheSerLysValLysThrLysThrProLysHis 361
Db 2311 AAGCAGAGCCAGGCGCTGCAGAACAGAGCTCGCATGCTCGGAGAGAGCTGCGGCGCA 2370
QY 362 SerProLleLysGluLupProCysGlySerLeu-----Ser 373
Db 2371 GGGCGTAGGACACCAAGCCCTCAAACTCCGTATTCCTCTCTGCTTCGAGACTCA 2430
QY 374 Glu-ThrValCysLysArgGluLeuArgSerGlnGluThrProGlu----- 388
Db 2431 GAGAAAGATTCTGCCAAGACCTTGCAATCTCAGAGAGGCCACAAGCATTGAGAGAG 2490
QY 389 -----LysProArgSerSerValAspThrProProArgLeuSerThrProGlnLys 405
Db 2491 CCAGATCTAGGCGCGAGGCGCAGCCAGCCTGCATGGGGGCTGTGTGCCAGAGCA 2550
QY 405 SerGlyProSerThrHisProLysGluLysAlaPhe----- 416
Db 2551 CCCACTGCCAACACAGCCTCTACAGAAAGCTTCTCTGCTAAGCCCGGCTCACACGCTG 2610
QY 417 ---SerSerGluIleGluAspLeuProTyrLeuSerThrThrGluMetTyrLeuCysAr 435
Db 2611 CGCCCCCGGAGCTTCCATCCCGACCAAGTGTCTCCGTCGACCTCGCTGATGTGGGC 2670
QY 435 gTP----- 436
Db 2671 CTGGGCGCGCAGGGCGCTGGGTGTGATGCTTGGCGGCTACTTTGTCAACAACTGTGCC 2730
QY 437 -HisGlnProProProSerProLeuProLeuArgGluSer-----Se 450
Db 2731 CCACAGCCCGCACCGCTGCGCCCGGCGGCTCTCCGCAAGCCCTGGAGATACAC 2790
QY 450 rProLysLysGluGluThrVal-----AlaArgCysLeuMetProSerSe 465
Db 2791 CCGAAACAGACACAGACACTGCTGATGAGAGGCTTCTGTGGTGGCGCGGCTTCAGGT 2850
QY 465 rValAlaGlyGluThrSerValLeuAlaValPro 476
Db 2851 GTCGGCGGGGCTGGCAGCGCGGTGTTGCTGCC 2884

Search completed: March 25, 2005, 21:32:06
Job time : 912 secs

Db 1036 CAGTTGAGGCCAAGAGAGAGATGAGAGACTGAATGCCAGAGAGATACGCTCTT 1095
 Qy 259 ALaArg1Ieg1uArgMetG1uArgMetG1uVal1uArg1uArg1uArg1uArg1u 278
 Db 1096 GCTCGAGTTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1155
 Qy 279 Arg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1u 298
 Db 1156 AGGCAAG 1215
 Qy 299 Lys1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1u 318
 Db 1216 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1275
 Qy 319 ProPheSerCysG1uArgSerG1uArgSerG1uArgSerG1uArgSerG1uArg 338
 Db 1276 CTTTCTCATGTGCGCGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1335
 Qy 339 GluArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1u 358
 Db 1336 GAAAG 1395
 Qy 359 ProLysHisSerProLysHisSerProLysHisSerProLysHisSerProLys 378
 Db 1396 CCAAG 1455
 Qy 379 ArgG1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1u 398
 Db 1456 CGTGAATTAAG 1515
 Qy 399 ArgLysSerThrProGlnLysG1uArgSerThrHisProLysG1uArgSerThr 418
 Db 1516 AGACTCTCGACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1575
 Qy 419 Glu1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg 438
 Db 1576 GAGATGAG 1635
 Qy 439 ProProSerProLysArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg 458
 Db 1636 CCTCCCAATCAAG 1695
 Qy 459 ArgCysLysMetProSerSerValAlaG1uArgSerValLysAlaValProSer 478
 Db 1696 AGGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1755
 Qy 479 ArgAspHisSerValGluProLysArgAspProLysArgAspProLysArg 498
 Db 1756 AGGGACCACTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1815
 Qy 499 ArgAspSerValPheSerLysArgHisAlaLysLysLysLysLysLysLys 518
 Db 1816 GATGACAGTGTATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1875
 Qy 519 ArgTyrAsp1Ieg1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg 538
 Db 1876 CGATGGATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1935
 Qy 539 TyrLysLysLysG1uArg1uArg1uArg1uArg1uArg1uArg1uArg1uArg 558
 Db 1936 TATTAAG 1995
 Qy 559 ArgAspValGluSerLysMet1IleThrProPheLysProValAlaPheGly 578
 Db 1996 GATGATGTAAAGTGTGCTGATTAACCTCTTCTGCTGCTGCTGCTGCTGCT 2055
 Qy 579 LysProLysLysThrProGlnLysPheGluLysLysLysLysLysLysLys 598
 Db 2056 TTGGCAAAATTAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2115
 Qy 599 ArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 614
 Db 2116 AGGTGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2163

RESULT 2 AK053719 LOCUS DEFINITION

AK053719 2839 bp mRNA linear HTC 03-APR-2004
Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched library, clone:EL130302B09 product:hypothetical protein, full insert sequence.

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AK053719 GI:26343664
AK053719.1 HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

JOURNAL

99279253
10349636

REFERENCE

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subcloning of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

20499374
11042159

REFERENCE

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, D., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multiplexed sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

20530913
11076861

REFERENCE

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL

5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL

6 (bases 1 to 2839)
Adachi, D., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shikama, A., Shikata, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>
Location/Qualifiers

FEATURES

source

```

1..2839
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:E130302B09"
/db_xref="taxon:10090"
/clone="E130302B09"
/tissue_type="eyeball"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
3'..1829
/note="unnamed protein product; hypothetical protein
(evidence: ProCrest)
putative"
/codon_start=1
/protein_id="BAC35489.1"
/db_xref="GI:26343665"
/translaction="MTMRSAVFRAAAPAGCNEDPRLDYERAAALGPEDESGAEAH
FLPRHRKLKEGPPLASQSGSPSPAGCGGKRGGLLPAGAAQOQESGSEVP
LCPRPATQAGIGGEPVAAAGAGSPRKTQAVLPQTGSIIVAAAKEPMPMGDGE
AAPATASDPAGPPLPLPGPPLPATNAGTLAASBEGMIRKSPGPGGSGAS
SOACIKOILLOLDLIEOQOOLQAKEXEIKELSEBRLILAIRMRMOLVYRD
NEKERHKLQGYPERBEREBEASEKIKLEBPCLCETSQLPKSPSCGSGGKHKK
TPPGNTERKTVPVKLAPEFSKVKTKPKSHPIKEPCGSISETVCRELRSQETPKP
RSSVDVTPRLSTPQKSGSTHPKEKASSEMEDLPYISTEMYCRMHQPPSPPLPBE
SSPKKEITVAIPMRDHSVEPLRPNPSDILENLDVSFSKRHAKLEDEKRRKWDI
ORIREORILORLOLRMYKKKGIQESSEPVTSFPPEPDVYSILITFLPVAAGRLP
KLAQPQELPMDENAPFLQVGPDEAHTSPDL"

```

CDS

```

33..1829
/note="unnamed protein product; hypothetical protein
(evidence: ProCrest)
putative"
/codon_start=1
/protein_id="BAC35489.1"
/db_xref="GI:26343665"
/translaction="MTMRSAVFRAAAPAGCNEDPRLDYERAAALGPEDESGAEAH
FLPRHRKLKEGPPLASQSGSPSPAGCGGKRGGLLPAGAAQOQESGSEVP
LCPRPATQAGIGGEPVAAAGAGSPRKTQAVLPQTGSIIVAAAKEPMPMGDGE
AAPATASDPAGPPLPLPGPPLPATNAGTLAASBEGMIRKSPGPGGSGAS
SOACIKOILLOLDLIEOQOOLQAKEXEIKELSEBRLILAIRMRMOLVYRD
NEKERHKLQGYPERBEREBEASEKIKLEBPCLCETSQLPKSPSCGSGGKHKK
TPPGNTERKTVPVKLAPEFSKVKTKPKSHPIKEPCGSISETVCRELRSQETPKP
RSSVDVTPRLSTPQKSGSTHPKEKASSEMEDLPYISTEMYCRMHQPPSPPLPBE
SSPKKEITVAIPMRDHSVEPLRPNPSDILENLDVSFSKRHAKLEDEKRRKWDI
ORIREORILORLOLRMYKKKGIQESSEPVTSFPPEPDVYSILITFLPVAAGRLP
KLAQPQELPMDENAPFLQVGPDEAHTSPDL"

```

ORIGIN

Alignment Scores:

Pred. No.: 1,52e-167 Length: 2839
Score: 3005.00 Matches: 575
Percent Similarity: 94.98% Conservative: 11
Best Local Similarity: 93.19% Mismatches: 12
Query Match: 92.80% Indels: 19
DB: Gaps: 3

US-10-054-935-2 (1-614) x AK053719 (1-2839)

```

QY 1 MetThMeCArgSerAlaValPheLYsAlaAlaAlaProAlaGlyGlyAenProGlu 20
DB 33 ATGACCATGAGATCCGCAATGTTCAAGGCGCGCGCCCTCGCGGCAACCCCGAG 92
QY 21 GlnArgLeuAspTyrGluArgAlaAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
DB 93 CACGAGCTGACCTACGAGCGGCGCTGCGCGCGCGCCAGAGCGAGTCCGGGGCG 152
QY 41 AlaGluAlaHisPheLeuProArgHisArgLYsLeuLYsGluProGlyProProLeuAla 60
DB 153 GCGGAAGCCATTTCCTCCCGCGCATCTTAAGCTCAAGAGCGGGGCCCGCTGGCC 212
QY 61 SerSerGlnGlySerProAlaProSerProAlaGlyCyS---GlyGlyLYsGlyArg 79
DB 213 TCTTCCAGGGCGGAGCGCTCGCCCTCTCCAGCGGCTGCGGCGGCGCAAGGCGCG 272
QY 80 GlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGln 99
DB 273 GGGCTGTGTATCTCCAGCGGGGCGGCGCGCGGAGAGGAAGAGACTGGGCGGTTCG 332
QY 100 ValProLeuProCySProProProAlaThrLYsGlnAlaGlyLYsGlyGlyGlyGly 119
DB 333 GTGCCCTTGCCCTGTCCGCGCCCAAGCTACTTAACCAAGCCGCGATCGCGGGGAGCCAGTC 392

```

```

QY 120 AlaAlaGlyAlaGlyCySerProArgProLYsTyrGlnAlaValLeuLeuProGlnThr 139
DB 393 GAGCGGGCGGTGGCGTGGCGCCCGCCCAAGTATCAAGCGGCTGTGCTCATTCAGAG 452
QY 140 GlySerLeu---ValAlaAlaLYsGluProThrProTyrAlaGlyAspLYsGlyGly 158
DB 453 GGGCTCATGCGTGGCGCGCGCCAAAGAGCTTACGCCCTGGGGCTGGGAGCAAGGTGGG 512
QY 159 AlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeuPro 178
DB 513 GCGGCTCCCCAGCTGCCACCCCTCTGAGACCCGCGGAGCCCAACCACTACTCTCGCCC 572
QY 179 GlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlnGlyArg 198
DB 573 GGGCCACCAACCCCTGCGCCACCGCACTGCTGGAGCCCTGAGCGGCAAGTGAAGGCA 632
QY 199 TTPLYSerMetArgLYsSerProLeuGlyGlyGlyGlyGlySerGlyAlaSerSerGln 218
DB 633 TGGAGAGTATTAAGAAAGGCCCTCTCGGGGGTGGCGGCGCTCGGAGCCCTCGACTAG 692
QY 219 AlaAlaCYsLeuLYsGlnLeuLeuLeuGlnLeuAspLeuIleGlnGlnGlnGln 238
DB 693 GCGGCTGCTCAACAGATCCTTCTGTGCAATTGAACTCATCGAAGCAGCAGCAGCAG 752
QY 239 GlnLeuGlnAlaLYsGlnLYsGlnIleGlnGlnLeuLYsSerGlnLYsAspThrLeu 258
DB 753 CAGTTGCAGCGCAAGAGAGAGAGATAGAGAGCTGAATCCGAGAGATAGCGCTCTT 812
QY 259 AlaArgIleGlyArgMetGluArgArgMetGlnLeuValLYsAspAsnGlyLYsGln 278
DB 813 GCTCGGATTTGAACGTTAGAAAGCGGATGCAAGCTGTGTAAAGAGATTAACAGAAAGAA 872
QY 279 ArgHisLYsLeuAspLeuGlnLYsTyrGlnThrGlnGlnGlnGlnGlnGlnGln 298
DB 873 AGGCACAACCTGCTCCAGGGCTATGAACCTGAAGAGAGAGAGAGCAAGATTGTCTGAG 932
QY 299 LYsIleLYsLeuGlnCYsGlnProGlnLeuSerGlnThrSerGlnThrLeuProLYs 318
DB 933 AAAATTAAATTGAGAGCGCAGCGGAGCTTTCGAACTCCCAAGCTGTGCTTCCAAAG 992
QY 319 ProPheSerCYsGlyArgSerGlyLYsGlyHisLYsArgLYsSerProPheGlySerThr 338
DB 993 CTTTTCATGTGGCGCGAGTGGAAAGGACACAAAGAAACCATTGGAAATATCA 1052
QY 339 GlnArgLYsThrProValLYsLYsLeuAlaProGlnPheSerLYsValLYsThrLYsThr 358
DB 1053 GAAAGAAAGACTCTGTAAAGCTGCTCTGAAATTTTCAAAAGTCAAAACAAAGACT 1112
QY 359 ProLYsHisSerProIleLYsGlnGlnProCYsGlySerLeuSerGlnThrValCYsLYs 378
DB 1113 CTTAAGCACTCTCCCATTTAAAGAGAAACCTGTGTGTTCAATATCAAGAACTGTTGTAA 1172
QY 379 ArgGluLeuArgSerGlnGlnThrProGlnLYsProArgSerSerValAspThrProPro 398
DB 1173 CTTGAATTAGAGAGCCAAAGAAACCCCAAGAAAGCCCGGTCTTCACTGATACCCACCA 1232
QY 399 ArgLeuSerThrProGlnLYsGlyProSerThrHisProLYsGlnLYsAlaPheSerSer 418
DB 1233 AAGCTCTCAAGTCCCAAAAGGAGCCACAGCAACCCCAAGAGAAAGACTTCTCAAGT 1292
QY 419 GlnIleGlnAspLeuProTyrLYsSerThrThrGlnMetTyrLeuCYsArgTyrPheGln 438
DB 1293 GAGATGGAAGATTGGCGTCACTTTCACACAGAAATATATTGTGTGGCGCACAG 1352
QY 439 ProProProSerProLeuProLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 458
DB 1353 CTTCCCAATCACCGTTACATTAACGGAATCTCTCCAAAGAGAGAGACTGTA--- 1409
QY 459 ArgCYsLeuMetProSerSerValAlaGlyGlnThrSerValLeuAlaValProSerTyr 478
DB 1410 -----GCATTTCTTCTTGG 1424
QY 479 ArgAspHisSerValGlnProLeuArgAspProAsnProSerAspLeuLeuGlnAsnLeu 498

```


QY	455	GIlnthrtValAlaAqCybLeumErProSerSerValAlaagiYglUnthSerValLeuAla	474
Db	866	GAGCTGTAGCAAGAGTGTCTAGTCCCAACAGTGTGGAGAGAAACTTCAGTCTTGCT	925
QY	475	ValProSerTTPAqASpHisSerValGluProLeuAqAaPProAsnProSerAspLeu	494
Db	926	GTTCTCTTCTTGGAGGACCACTCAGTAGAGCTCTTAAGGAGCCAAACCTTCAGACCTT	985
QY	495	LeuGluLeuLeuAaPAspSerValPheSerLySarGHisAlaLybLeuGluLeuAaPglu	514
Db	986	TTGGAGAACCTCGATGACAGTGTGTTTCGAAGCGGCATGCAAAACTGGACCTGGATTAG	1045
QY	515	LySarTgrtLybSAqTTPAspIIGlnArgIIGIearGluGlnAqGILeLeuGlnArgLeu	534
Db	1046	AAGGAGAGAAAGAAAGTGGATATTTCAGAGCATCAGGAGACAAAGATTTCACGCACTG	1105
QY	535	GIlnLeuAqNectITyLybLybLeGlyLIGlnGluSerGluProGluValThSerPhe	554
Db	1106	CAGCTCAGAAATGTTAAATAAAGAAAAGAAATTCAGAGATCTGAGCCTGAGATTCTCATTT	1165
QY	555	PheProGluProAaPAspValGluSerLeuNectIETrrProPheLeuProValValAla	574
Db	1166	TTCCCTGAGCAAGATGATGTTGAAAGTTTGAATGATTACCCTTCTTGCTGTGTAGCA	1225
QY	575	PheGlyAqProLeuProLybLeuThrProGlnAsnPheGluLeuProTyrPLeuAaPglu	594
Db	1226	TTTGGACGACCATTAACCAAAATTAACTCCACAGAAATTTTGAGTAACTCGTGATGATAG	1285
QY	595	ArgSerArgCybAqLeuGluIIGlnLybLybGlnThrProHisSarGThCybArgLyb	614
Db	1286	CGTAGCCGATGCAAGTTGGAGATCCAGAAAGACAAACACTTCACCGAGCTGTAGGAAA	1345

RESULT 4	
AKO15496	
LOCUS	AKO15496 1304 bp mRNA linear HTC 03-APR-2004
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930463f05 product:hypothetical protein, full insert sequence.
ACCESSION	AKO15496
VERSION	AKO15496.1 GI:12853861
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1
AUTHORS	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	2049374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, U., Nishi, K., Kitenuma, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kiru, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipicillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861

REFERENCE	TITLE	JOURNAL	AUTHORS
4	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
5	Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)		
REFERENCE	TITLE	JOURNAL	AUTHORS
5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)		
	(bases 1 to 1304)		
	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Ari, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukushima, Y., Furumori, M., Hanagaki, T., Hara, A., Hayatsu, N., Himemoto, K., Hirooka, T., Horii, F., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kesukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kuwahara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okio, T., Owa, C., Satou, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Saeki, D., Shibata, K., Shihata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.		
	Direct Submission		
	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, url:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)		
	Please visit our web site (http://genome.gsc.riken.jp/) for further details.		
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAAGACATCCAAAGCCTCTTTTCTTTTTNN 3']. cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAAGATTCGCAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size complement longer than 7 kb was selected before cloning. Vector: a modified plasmid pBluescript KS(+) after bulk excision from lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.		
	Location:Qualifiers		
	1. .1304		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="CS7BL/6J"		
	/db_xref="FANTOM.DB:4930463F05"		
	/db_xref="taxon:10090"		
	/clone="4930463F05"		
	/sex="male"		
	/tissue_type="testis"		
	/clone_lib="RIKEN full-length enriched mouse cDNA library"		
	/dev_stage="adult"		
	17. .1129		
	/note="unnamed protein product; hypothetical protein (evidence: ProCrEst)"		
	putative"		
	/codon_start=1		
	/protein_id="BAB29868.1"		
	/db_xref="GI:12853862"		
	/translation="MLRFAHGOEALIPSLAQTTTNKRLAIRERMRMQLVYK RDNKEHKHLQGVPEPERAEASEIKTKRGSPICETSOALPSKPFGSGSKGRK RPCKNGTEKTPVKTLAPFKAFKTKTGPHSIIKPEPCOSISETVCKRKLRSQEPD KPSSVDTPRLSTPQKGSPTRPKKKAFFSEMEDLPILSTTEWTLCNHDPPSPPLP LRESSPKRTVALPSWMDVEVLRPNDPDIILENDSDVSFRNALVELDEKRKKW DIRIRPRKORTLQRLMTMYKKKGIQSEBPVTSFEEDPVESLIPLPVPVAFGP		

US-10-054-935-2 (1-614) x BU857520 (1-927)

```
QY 271 ValIyelaSpaSpangIuIySguAqrGHiLeIyLeuPheGlnGlyYrGluThrGluGlu 230
Db 9 GTAAAGAGGATTAACGAGAAAGAAAGCAACAGCTGTTTCAGGGCTATGAACCTGAAGAG 68
QY 291 ArgGluGluThrGluLeuSerGluYsIleIyLeuGluGluGlnProGluLeuSerGlu 310
Db 69 AGAGAGGAAACAGAGCTATCTGAGAAATTAATTAACCTGAGAGCCAGCGAGCTTCCGAG 128
QY 311 ThrSerGlnThrLeuProProLysProPheSerCysGlyYarSerGlyYsGlyYsIle 330
Db 129 ACATCCAGACTGCTGCTCCCAAGCCCTTCTCATGtGGGAGTGGAAAGGAGCATAAA 188
QY 331 ArgLysSerProPheGlySerThrGluYarGlyYsThrProValIySleuAlaProGlu 350
Db 189 AGGAAATCCCATTTGGAAAGTACAGAAAGAAAGAAAGACTCTGTAAAGAGCTGCTCCGAA 248
QY 351 PheSerLysValIySthrIySthrProLysHisSerProIleIySgluGluProCysGly 370
Db 249 TTTTCAAAAGTCAAAACMAAAACTCTTAACACTCTCTTAAGAGAAACCTGTGCT 308
QY 371 SerLeuSerGlnThrValCysLysArgGluLeuArgSerGlnGluThrProGluLysPro 390
Db 309 TCCTTATCTGAACCTGTTTGTAAACGTAATTGAGAGCCAGAAACCCAGAAAGCC 368
QY 391 ArgSerSerValAspThrProProArgLeuSerThrProGlnYsGlyProSerThrHis 410
Db 369 CGGTTTCAGATGACACCCACCAAGACTCTCCACTCCCAAAAGGAGCCAGACCAT 428
QY 411 ProLysGluLysAlaPheSerSerGluIleGluAspLeuProTyLeuSerThrGlu 430
Db 429 CCAGAGAGAAACCTTCTCAAGTGAATGAAGAAATTTGCCGTAACCTTTCACACAGAA 488
QY 431 MetTyLeuCysArgTrpHisGlnProProSerProLeuProLeuArgGluSerSer 450
Db 489 ATGTAATTTGTGCTTGACACCAAGCTCCCTCCATCACCGTTACATTACGGAACTCT 548
QY 451 ProLysGluGluThrValAlaArgCysLeuMetProSerSerValAlaGlyGluThr 470
Db 549 CCAGAGAGAGAGAGACTGTAGCAAGGTGTCTGATGTCATCAAGTGTGACAGAGAACT 608
QY 471 SerValLeuAlaValProSerTrpArgAspHisSerValGluProLeuArgAspProAsn 490
Db 609 TCAGTCTTGCTGCTTCTCTTGGAGGAGACACTCACTAGAGCTCTMAAGGACCCAAAT 668
QY 491 ProSerAspLeuLeuGluAsnLeuAspAspSerValPheSerLysArgHisAlaIySleu 510
Db 669 CCTTCAGACCTTTTGGAGAACCTGGATGACAGTGTGTTTTCGAAAGCGGATGCAAACTG 728
QY 511 GluLeuAspGluLysArgArgLysArgTrpAspIleGlnArgIleArgGluGln-ArgIle 530
Db 729 GAGCTGGATGAGAAAGAAAGAAAGAAAGTGGATATTCAGAGGATCAGCGGAAACAAAGAT 788
QY 530 eLeuGlnArgLeuGlnLeuArgMetTyTyLysLysLysGlyIleGlnGluSer--GluPr 549
Db 789 TTAACGACCGACTCAGCTCAGAAATGTAATAAAAGAGGATTTTCAGAAATCTGAGCC 848
QY 549 ogIu-ValThrsSerPhePro 556
Db 849 TGAAGGTACTCAATTTTTC 871
```

RESULT 6
BX329236/c
LOCUS BX329236 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION CDNA clone CS0DK002YN11 3-PRIME, mRNA sequence.
ACCESSION BX329236
VERSION BX329236.2 GI:46280783
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 896)
Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 2, 2003 this sequence version replaced gi:30344795.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
8989.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS0BA1036ZB04_CS03386_1&c=8989.f

FEATURES
source

Location/Qualifiers
1..896
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK002YN11"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 2,026-71 Length: 896
Score: 1370.50 Matches: 279
Percent Similarity: 95.89% Conservative: 1
Best Local Similarity: 95.55% Mismatches: 12
Query Match: 42.33% Indels: 4
DB: 5 Gaps: 0

US-10-054-935-2 (1-614) x BX329236 (1-896)

```
QY 77 LysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGluSerTrp 96
Db 894 CAAGGCGCGGGGCTGTACT-CCGNNCGGGGGGCGCCCGGGGAGCAGAGAGAGCTGG 836
QY 97 GlyGlySerValProLeuProCysProProAlaThrLysGlnAlaGlyIleGlyGly 116
Db 835 GNCGGTCG-GTGGCCCTGCTGTCGCC-CCGGGCACCAAGCAGCGGGCAT-GGGGGG 779
QY 117 GluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysThrGlnAlaValLeuPro 136
Db 778 GAGCTCGCNGCGAGCGGAGCGGGCTGCACCCCGGCGCAAGTATGAGCGGTGCCCC 719
QY 137 IleGlnThrGlySerLeuValAlaAlaAlaLysGluProThrProTrpAlaGlyAspLys 156
Db 718 ATTCAACGGGCTCTCTGCGCGNGCGCCAAAGAGCTACGCCCTGGGCTGNGAGCAAG 659
QY 157 GlyIyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProLeuPro 176
Db 658 GGTGGGGCGGCTCCNCCGCTGCCACCGCTCGGAGCCCGGGGAGCCCCACACATACT 599
QY 177 LeuProGlyProProProLeuAlaProProAlaThrAlaGlyThrLeuAlaAlaSerGlu 196
Db 598 CTGCCCGGGCGCCACCTCTCGCGCCACCGCCGAGCCTTGCGCGCCAGCGAG 539
QY 197 GlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAlaSer 216
Db 538 GGCAGATGAGAGATATGAGAAAGACCTCTCGGGGGGTGGTGGCGCTCGGAGGCTCC 479
```

OY		217	SerGlnAlaAalCyLeuLysGlnIleLeuLeuGlnLeuAspLeuIIeGIuGlnIn	236
Dd		478	AGTCAAGCCCGCTGCCTCCTAAACAGATCCTTCGTCTGCAATTGGACTCATGGAACGAG	419
OY		237	GInGlnGlnLeuGlnAlaLyseGluLysGluIleGlnGlnLeuLysSerGluArqApThr	256
Dd		418	CAGCAGCAGCTGCAGGCCAACGAAAAAGAGATCGAGGAGCTGAATCGAGAGAGACAAG	359
OY		257	LeuLeuAlaArgIleGluArqMetGluArqArqMetGlnLeuValLysLysAspSngLu	276
Dd		358	CTCCTGCTCGGATTGAACGTATGAAAAGCGGAGCGAGCTGGTTMAAGAAGATTACGAG	299
OY		277	LysGluArqGHIslvLysLeuPheGlnGlyrGrIuThrGluIuArqGluIuThrGluLeu	296
Dd		298	AAAGAAAAGCACAAAGCTGTTTCAGGGCTATGAATACTGAAGAGAGAGAGAAAACAGACTA	239
OY		297	SerGluVslIleLysLeuGluICySeGlnProGlnLeuSerGluThrSerGlnThrLeuPro	316
Dd		238	TCTGAGAAAATTAACTGGAAGTCCAGCCGAGAGCTTTCGAGACATCCCAAGCTTGCT	179
OY		317	ProlYsPropHesCySeGlyArqSerGlyLysGlyslvLysArgLysSerProBheGly	336
Dd		178	CCCAAGCCCTTCTCATGTGGGGGAGTGAAAGGAGACATAAAGAAATCCCATTTGGA	119
OY		337	SerThrGluArqLysThrProValIlyslvLysLeuAlaP-roGluPheSerLysValylsThr	356
Dd		118	AGTACAGAAAAGAAAGACTCTGTTAAAAAGCTGGCTCTGTGAATTTCAAAGTCMAAAC	59
OY		357	LysThrProlYsSHisSerProIleLysGlnIuPro	368
Dd		58	AAAACTCTTAAGCACTCTCTTAITTAAGAGAACCC	23
RESULT 7				
Bu612708				
LOCUS			773 bp	mRNA linear EST 20-FEB-2003
DEFINITION			UT-M-ER0-cbc-d-14-0-U r1 NIH_BMAP_PRO Mus musculus CDNA clone	
ACCESSION			BU612708	
VERSION			BU612708.1	GI:23278923
KEYWORDS			EST.	
SOURCE			Mus musculus (house mouse)	
ORGANISM			Mus musculus	
REFERENCE			Eukarya; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;	
AUTHORS			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE			1 (bases 1 to 773)	
JOURNAL			NIH-MGC http://mgc.nci.nih.gov/.	
COMMENT			National Institutes of Health, Mammalian Gene Collection (MGC)	
			Unpublished (1999)	
			Contact: Robert Stransberg, Ph.D.	
			Email: cga@bs-iemail.nih.gov	
			Tissue Procurement: Dr. Jim Lin, University of Iowa	
			cDNA library preparation: Dr. M. Bento Soares, University of Iowa	
			DNA Sequencing by: Dr. M. Bento Soares, University of Iowa	
			Clone Distribution: Clone distribution information can be obtained	
			from Dr. M. Bento Soares, bento-soares@iowa.edu	
			This clone was contributed by the Brain Molecular Anatomy Project	
			(BMAP)	
			Seq primer: DYK-5.	

```

FEATURES
source
location/Qualifiers
1..773
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="UI-M-FR0-cbc-d-14-0-UI"
/tissue_type="whole brain"
/dev_stage="embryo 13.5, 14.5, 16.5, 17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="NH BMAP FR0"
/note="Organ: Brain, Vector: pYX-Asc, Site: 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

```

1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCAGACAGC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., Program coordinator."

ORIGIN	Alignment Scores:	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
	4.17e-69	Length:	773	1330.00	Matches:	249	
				99.61%	Conservative:	1	
				96.89%	Mismatches:	7	
	41.07%				Indels:	0	
	5	Gaps:	0				
US-10-054-935-2 (1-614) x BU612708 (1-773)							
QY	323	GLYARSGERGLYSGLYSHLYSARYLSERProPheGISErThrGluArgLYSThr	342				
Db	3	GGCGGAGTGGAGAAAGGACACAAAAGAAAACCCATTGGAAATACAGAAAGAAAGACT	62				
QY	343	ProVALysLysLeuAlaProGluPheSerLYSValLYSThrLYSThrProLYSHISer	362				
Db	63	CCGTGTAAGAAAGCTGGCTCTGATTTTCAAAGTCAAAACAAACCTCTAACACACT	122				
QY	363	ProTLEYSGLUGLUPROCYSGLYSERLeuSerGluThrValCYLSYsArgGluLeuArg	382				
Db	123	CCCATTAAGAGGAACCTGTGTTCCATATCAGAAACTGTTTGTAAACGTGAATTGAGG	182				
QY	383	SeRGLInGluThrProGluLYSPROArgSERSerValAspThrProProArgLeuSerThr	402				
Db	183	AGCCMAAGAACCCAGAAAAGCCCCGGCTTCAGTGAATACCCCAAGACCTTCGACT	242				
QY	403	ProGlnLYSGLYProSerThRHisProLYSGluLYsAlaPheSerSerGluILEGLUAsp	422				
Db	243	CCCCAAAGAGGACCCGACACCCACCCAGAGAGAAACCTTCTCAAGTGAATGAAGAT	302				
QY	423	LeuProLYrLeuSerThrThrGluMetLYrLeuCYsArgTrpHisGlnProProSer	442				
Db	303	TTGGCCGACCTTCCACACAGAAATGTATTGTGTGTTGGACCAAGCCTCCCCATCA	362				
QY	443	ProLeuProLeuArgGluSerSerProLYsGLUGluThrValAlaArgCYsLeuMet	462				
Db	363	CCGTATACATTAGCGGAGATCCTCTCCAAAGAGAGAGACTGATGACAGAGTCTCATG	422				
QY	463	ProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTPARgAspHisSer	482				
Db	423	CCATCAAGATGTTCAGAGAGAACTCAGCTTGCGTGTCTCTTGGAGGACACTCA	482				
QY	483	ValGluProLeuArgAspProAspProSerAspLeuLeuGluLeuAspAspSerVal	502				
Db	483	GTPAGAGCTTAAAGGACCCAAATCTTTCAGACTTTTGGAGAACTCGATGACAGTGA	542				
QY	503	PheSerLYsArgHisAlaLYsLeuGluLeuAspGluLYsArgLYsArgLYsArgTPAspILE	522				
Db	543	TTTTCAAGAGGAGCATGCMAAACTGAGAGCTGACGAGAAAGAGAGAAACATGGGATATT	602				
QY	523	GluArgLYLeArgGluGluArgLYILEuGluArgLYGluLeuArgMetLYrLYsLYsLYs	542				
Db	603	CAGAGGATCGAGGAACNAGAAATTTTACACGACTGAGCTCGAAATGTTAAAAAGAAA	662				
QY	543	GlyLYleGluLeuSerGluProGluValThrSerPhePheProGluProAspAspValGln	562				
Db	663	GGAAITTCAGGAATCTGAGCTGAGGTTACTCATTTTCCCTAGACCAAGATGATGTTGAC	722				
QY	563	SeRLeuMetILEThrProPheLeuProValAlaPheGlyArgProLeu	579				

|||||:|||||
DB 723 AGTTTCTGATTACCCCTTCTTGCTTGTGACATTGGACGGCCATTG 773

RESULT 8
BQ014717/c 765 bp mRNA linear EST 26-MAR-2002
LOCUS UI-H-ED1-axe-j-04-0-UI.s1 NC1 CGAP_ED1 Homo sapiens cDNA clone
DEFINITION IMAGE:5833443 3', mRNA sequence.

ACCESSION BQ014717
VERSION BQ014717.1 GI:19739618
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 765)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TUMOR Tumor
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Seg primer: M13 FORWARD
POLYA=Yes.

FEATURES
Source Location/Qualifiers
1..765
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5833443"
/issue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI-CGAP_ED1"
/note="Organ: Left Pubic Bone; Vector: pT73-Pac
(Pharmacia) with a modified polylinker; Site 1: EcoR I;
Site 2: Not I; NCI CGAP ED1 is a normalized cDNA library
containing the following tissue(s): Chondrosarcoma cell
line C5. The library was constructed according to
Bonaldio, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into pT73-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is GCTCAGGCT.
TAG_TISSUE=chondrosarcoma
TAG_LIB=UI-H-ED1
TAG_SEQ=CCTCAGGCT"

ORIGIN
Alignment Scores:
Pred. No.: 9.31e-69 Length: 765
Score: 1324.00 Matches: 251
Percent Similarity: 99.60% Conservative: 0
Best Local Similarity: 99.60% Mismatches: 1
Query Match: 40.89% Indels: 0
DB: 5 Gaps: 0
US-10-054-935-2 (1-614) x BQ014717 (1-765)
QY 291 ArgGlnGluThrGluLeuSerGluTyrIleValLeuGluCysGlnProGluLeuSerGlu 310
DB 763 AGAGAGAAACAAGCATCTGAGAAATTAACCTGAGAGGCCAGCCGAGCTTTCCGAG 704

QY 311 ThrSerGlnThrLeuProGluProPheSerCysGluArgSerGluValGlyHisIle 330
DB 703 ACATCCCAAGCTCTGCTCTCCAAAGCCCTTCTCATGTGGCGAGGTGAAGACATTA 644

QY 331 ArgIleSerProPheGlySerThrGluArgGlySerProValIleValLeuAlaProGlu 350
DB 643 AGGAATCCCACTTTGGAAGTACAGAAAGAAAGACTCCTGTAAAGCTGGCTCTGAA 584

QY 351 PheSerValIleValThrIleThrProGlyHisIleSerProIleValGluGluProCysGly 370
DB 583 TTTTCAAAAGTCAAAACAAAACCTCTTAAGCACTCTCCATTAAGAGAGAACCTGTGGT 524

QY 371 SerLeuSerGluThrValCysIleArgGluLeuArgSerGluGluThrProGluIlePro 390
DB 523 TCTTATCTGAACCTTTGTGTAACGTGAATTGAGAGCCAGAAACCCAGAGAAAGCCC 464

QY 391 ArgSerSerValAspThrProProArgLeuSerThrProGluIleValGlyProSerThrHis 410
DB 463 CGGTCTTCAGTGAGACCCCAACCAAGACTCTCACTCCCAAAAGGAGCCAGCAACCAT 404

QY 411 ProlYleGluValAlaPheSerSerGluIleGluAspLeuProTyrLeuSerThrThrGlu 430
DB 403 CCCAAGAGAAAGCCCTTCAAGTGATGAGAAAGATTGCCGTAACCTTCCACACAGAA 344

QY 431 MetTyrLeuCysArgTrpHisGlnProProProSerProLeuProLeuArgIleSerSer 450
DB 343 AGTATTTGTGCTGTGGACCAAGCTCCCAATCACTGATACATTAGCGAATCTCT 284

QY 451 ProlYleGluGluThrValAlaArgCysLeuMetProSerSerValAlaGlyIleThr 470
DB 283 CCAAAAGAGAGAGACTTACCAAGGTCTGATGCCATCAAGTTCAGAGAGAACT 224

QY 471 SerValLeuAlaValProSerTrpArgAspHisSerValGluProLeuArgAspProAsn 490
DB 223 TCAGTCTTGCTGTTCTCTTCTTGGAGGACCACTCAGTAGAGCTCTTAAGGAGCCCAAT 164

QY 491 ProSerAspLeuLeuGluLeuLeuAspAspSerValPheSerTyrArgHisAlaValLeu 510
DB 163 CCTTCAGACCTTTTGGAGAACCTGGATGACAGTGTCTTTCGAGGGCATGGAATACG 104

QY 511 GluLeuAspGluValysArgArgIlePhePheLeuGluArgIleArgGluGluArgIle 530
DB 103 GAGCTGATGAGAGAGAGAGAGAGAGATTTCAAGATTCAGAGATTCAGGAGACAAAT 44

QY 531 LeuGlnArgLeuGluLeuArgMetTyrIleValLeu 542
DB 43 TTACAGCGACTGACGCTCAGAAATGATTAATAAAAAA 8

RESULT 9
BQ229745 870 bp mRNA linear EST 02-MAY-2002
LOCUS AGENCOURT 7510022 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6042882
DEFINITION 5', mRNA sequence.
ACCESSION BQ229745
VERSION BQ229745.1 GI:20411145
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 870)
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.
TUMOR Tumor
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LHAM13282 row: p column: 19
High quality sequence stop: 596.

FEATURES

source

1. 870
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6042882"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."

ORIGIN

Alignment Scores:

Pred. No.: 2,22e-67 Length: 870
Score: 1302.00 Matches: 260
Percent Similarity: 92.07% Conservative: 7
Best Local Similarity: 89.66% Mismatches: 14
Query Match: 40.21% Indels: 9
DB: 5 Gaps: 1

US-10-054-935-2 (1-614) x BQ229745 (1-870)

QY 13 ALAPROIALAGLYGIAENPFGIUGINAGYLEUASPTFGIUGALAAALALEGLY 32
DB 2 GCCCTGCGCGCGCAATCTGAGACGACTGACACGAGCGGCGCTGCGCTGCGC 61
QY 33 GYIYPRGILUASPTGIPROGILYALAAALAGLUAALAHISPHLEUPROARGHISARYLEU 52
DB 62 GGGCCCGAGAGAGAGCTGGGGGGGCGGAGCCGACCTTCCCGCGCACCGTAACCTC 121
QY 53 IYSGIUPROGILYPROLEUALASERSEGINLYGISERPROALAPROSERPROA 72
DB 122 AAGGAGCGGGGGCGCGCTGCTCTCCAGGCGGAGCGCGCGCTTCCCGCGC 181
QY 73 GYIYSGIYGLYLYSGIYATGGLYVLEUULEUPROALAGLYAALAPROGILY 92
DB 182 GCGTGGCGCGGAGGCGCGGCGCTTGTACTCCGCGCGGCGCGCGCGCGGAG 241
QY 93 GUGIUSERTRPGIYGLYSEYVALPROLEUPROCYSPROPROALATHRYGALA 112
DB 242 GAAGAGAGCTGGGGGGGTGCGTCCCTTGGCTGTCGCGCGCGCACCAAGAGCC 301
QY 113 GYIYIYGLYGLYUPLROALAAALAGLYAAGLYCYSSERPROATGPROLYTYRGLN 132
DB 302 GGCATTGGGGGGGAGCCTGCGCGAGCGGAGCGGCTGCAAGCCCGCGCGCAATACAG 361
QY 133 ALVALLEUPROILEGINTHRYGISEYLEUVALAAALAAIYSGIUPROTHRTIP 152
DB 362 GCGGTGCTGCATTACAGAGGCGCTCTGCTGGCGGCGGCAAGAGCTTACGCGCTGG 421
QY 153 ALGLYASPIYSGIYGLYALAAASERPROALAAALHRAASERAPPROALAGLYPRO 172
DB 422 GCTGGGAGACAGGGTGGGGCGGCTCCCGCGTGCACCGGCTGAGCCGCGGAGACC 481
QY 173 PROPROLEUPROLEUPROGILYPROPROPROLEUALAPROTHRALATHRALAGLYTHLEU 192
DB 482 CCACCACTACTCTGCGCGCGCGCGCGCACCTTCCGCGCGCACCGCGGAGACCTTG 541
QY 193 ALAALASERGLUGLYATGTPLYSERMETARYLYSERPROLEUGLYGLYGLYGLY 212
DB 542 GCGGCGGAGAGGAGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 601
QY 213 SERGLYASERSEYGINALAAALCYALEUYSGLINLEUULEUAGINLEUASPLEU 232
DB 602 TCGGAGAGCTTCCATCTGAGCGCGCTGCTCAAGAGATCTTCTGCTGCAATTTGAGCTTC 661
QY 233 ILEGIUGINGLINGINGLINGINLEUGINALALYSGIUPRSGIUNLEGIUGIULEYSEY 252

DB 662 ATGAACAGACAGACAGAACAGCTGAGCGGCGCAAGAAAAGAGATCGAGAGCTGAAGTCA 721
QY 253 -GIUARGSPTRHLEULEU-ALARGILEGU-ARGMEGIUARGAGMETGILEUVAL 271
DB 722 NAAAGAAAACCTCTTCTGCTCGATTGAAAGATGAGAGAGCGGAGATGACAGCTCG 781
QY 272 IYLYSPASPSANGLUYSGLIARGHISLYLEUPHGINGLYTYR----- 286
DB 782 GTAAAGAGAGATACGAGAAAGAGAGAGGAGACACCAAGCCTGTTTCAAGGCGCTA 841
QY 287 ---GIUThrGluGluArgGluGlu 293
DB 842 TTAAGAACTTTGAAAAAGAGAA 865

RESULT 10

LOCUS

CK635536 769 bp mRNA linear EST 28-JAN-2004
DEFINITION U1-M-HNO-cl-m-20-0-UI.r1 NIH_BMAP_HNO Mus musculus CDNA clone
IMAGE:30643411 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CK635536 769 bp mRNA linear EST 28-JAN-2004
U1-M-HNO-cl-m-20-0-UI.r1 NIH_BMAP_HNO Mus musculus CDNA clone
IMAGE:30643411 5', mRNA sequence.
CK635536
CK635536.1 GI:41361402
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 769)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Straube, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.iowa.edu/distribution/mousefl.html>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: PYX-5.
Location/Qualifiers

FEATURES

source

1. 769
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30643411"
/tissue_type="Upper Head"
/dev_stage="9.5-10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_HNO"
/note="Organ: Head; Vector: PYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GAACTGAGT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:

Pred. No.: 8.83e-64 Length: 769
Score: 1240.00 Matches: 239
Percent Similarity: 95.31% Conservative: 5

Best Local Similarity:	93.36%	Mismatches:	12
Query Match:	38.30%	Indels:	1
DB:	7	Gaps:	0

[illegible]

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Daniela S. Gerhard, Ph.D.

Source

```

/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone_image="7126489"
/clone_type="kidney, pooled"
/lab_host="DH10B TconA"
/clone_id="NIH MGC 236"
/note="Organ: kidney; Vector: pXpress-1; Site 1: EcoRV;
Site 2: NotI; RNA obtained from pooled kidney tissue from
a mix of male and female animals at 8 wk old. Tissues were
snap-frozen and kept at -80C for two days before RNA
extraction and purification (TRI-reagent method). cDNA was
primed using oligo-dT primer:
5'-pGACTAGTCTAGATCGAGCGGCCGCC(CT)25-3' and cloned into
the EcoRV/NotI sites of pXpress-1. Size selection >1.4kb
resulted in an average insert size of 2.2 kb. This primary
library is normalized (non-normalized primary library is
NIH MGC 235) and was constructed by Express Genomics
(Frederick, MD). Note: this is a NIH MGC library."

```

ORIGIN

Alignment Scores:

Pred. No.:	1.73e-63	Length:	76
Score:	1235.00	Matches:	23
Percent Similarity:	96.84%	Conservative:	7
Best Local Similarity:	94.07%	Mismatches:	8
Query Match:	38.14%	Indels:	2
DB:	7	Gaps:	0

US-10-054-935-2 (1-614) x CK471650 (1-762)

Qy	270	GLuLYSGluARgHISylsleuPheGlnGlyrgrGluThrGluuAArgGluGluThrGlu	259
Db	2	GAAGAGAAAGGACAAAGCTGTTTCAGAGGCTATGAACTGTAAGAGAGAGAGAAACGAG	61
Qy	296	LeuSerGluLYsleuLeuGluCysGlnProGluLeuSerGluThrSerGlnThleu	315
Db	62	TTGTCTGAGAAATTTAACTGAGAGGCCAGCCGAGCTCTCGAGACATCTCCAGGCTCG	121
Qy	316	ProProLYsPProPheSerCysGlyAArgSerGlyLYsGlyHISylsAArgLYsSerProPhe	335
Db	122	CTCTTCAAGCCTTTCTCATGTGGCCGAGATGGCAAGACACAAAGAAACCCCATTT	181
Qy	336	GlySerThrGluArgLYsThrProValLYsLYsleuAlaProGluPheSerLYsValLYs	355
Db	182	GGAAATACAGAAAGAAAGACTCTGTAAAAAGCGGCTCTGTAATTTTCAAAATTCAA	241
Qy	356	ThrLYsThrProLYsHISerProIleLYsGluGluProCysGlySerLeuSerGluThr	375
Db	242	ACAAAACCTCTAAGACTCTCCCATTTAAAGAGAGCCCTGTGGTTCATTCAGAACT	301
Qy	376	ValCysLYsArgGluLeuAArgSerGlnGluThrProGluLYsProAArgSerValAsp	395
Db	302	GTTTGTAAACGTGAATTTGAGAGCCAAAGAACCCCAAGAAAGCCCGGTCTTCAGTGAT	361

QY 396 ThrProArgLeuSerThrProGlnLysGlyProSerThrHisProLysGluLysAla 415
 DB ACCCCACCAAGACTCTCGACTCCCAAAAAGGACCCCAAGCCCAAGAGCAAGGCC 421
 QY 416 PheSerSerGluIleGluAspLeuProTyrLeuSerThrGluMetTyrLeuCysArg 435
 DB 422 TTCTCAAGTAGAATGGAAATTTGGCTTACCTTTCACCAAGAAATGATTTGTGTGT 481
 QY 436 TrpHisGlnProProSerProLeuProLeuArgGluSerSerProLysGluGlu 455
 DB 482 TGCCACACAGCTCCCAATACCGTTTACATTACGGGAATCTCTCCAAAGAGAGAG 541
 QY 456 ThrValAlaArgCysLeuMetProSerSerValAlaGluGluThrSerValLeuAlaVal 475
 DB 542 ACTGTGACAAAGTGTGTGATGATCCATCAAGTGTGCAGAGAAACTTCACTTGCTGT 601
 QY 476 ProSerThrArgSerHisSerValGluProLeuArgSerProAsnProSerAspLeu 495
 DB 602 CTTTCTTGGAGGACCACTAGTAGACCTCTAAGGAGCCCAATCTTGACATTTTG 661
 QY 496 GluAsnLeuAspSerSerValPheSerLysArgHisAlaLysLeuGluLeuAspGluLys 515
 DB 662 GAAACCTCGATACAGTGTGTCTTTCAAAGAGCATCGAAACTGAGCTAGACGAGAG 721
 QY 516 ArgArgLysArgTTPAspIleGlnArgGluGluGln 528
 DB 722 AGAG-AAAAGATGGATAT-CAGAAATCAGGGAACA 758
 RESULT 12
 CN355143 743 bp mRNA linear EST 16-MAY-2004
 LOCUS 1700600188342 GRN_PRENU Homo sapiens cDNA 5', mRNA sequence.
 CN355143
 VERSION CN355143.1 GI:47355077
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 743)
 Brandenberger, R., Wei, H., Zhang, S., Lei, S., Muraige, J., Fisk, G., J.,
 Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,
 Lebkowski, J. and Stanton, L.W.
 Transcriptome characterization elucidates signaling networks that
 control human ES cell growth and differentiation
 Nat. Biotechnol. 22 (6), 707-716 (2004)
 CONTACT: Brandenberger R
 Regenerative Medicine
 Genon Corporation
 230 Constitution Drive, Menlo Park, CA 94025, USA
 Tel: 650 473 8658
 Fax: 650 473 7760
 Email: rbrandenberger@genon.com
 Insert length: 743 Std Error: 0.00.
 Location/Qualifiers
 1..743
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic stem cell, retinoic acid and
 mitogen-treated hES cell line H7"
 /clone_lib="GRN PRENU"
 /note="oligo dT primed, full-length enriched cDNA library
 from hES cell line H7 (p29) maintained in feeder-free
 conditions. Embryoid bodies were generated in the presence
 of all-trans retinoic acid and mitogens."

ORIGIN
 Alignment Scores: 1.69e-62 Length: 743
 Score: 1218.00 Matches: 234
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 37.62% Indels: 0
 DB: 7 Gaps: 0
 US-10-054-935-2 (1-614) x CN355143 (1-743)
 QY 225 IleLeuLeuLeuGlnLeuAspLeuIleGluGlnGlnGlnGlnGlnAlaLysGlu 244
 DB 29 ATCTTCTGTGTGAATTTGACCTCATCGAACACACACAGAGAGCTGCGCCAGGAA 88
 QY 245 LysGluIleGluGluLeuLysSerGluArgAspThrLeuLeuAlaArgIleGluArgMet 264
 DB 89 AAGGAGATGAGAGCTGAAGTACAGAGAGACACCGCTCTTCTCGATTGAACGATG 148
 QY 265 GluArgArgMetGluLeuValLysLysAspAsnGluLysGluArgHisLysLeuPheGln 284
 DB 149 GAAAGCGGATGACGTGGTAAAGAAAGATTAAGAAAGAAAGCAAGCTGTTTCA 208
 QY 285 GlyTyrGluThrGluGluLysGluGluGluGluGluGluGluGluGluGluGluGlu 304
 DB 209 GGCTATGAAACTGAAAGAGAGAGAGAAACAGACTATCTGAGAAATTTAAACTGAGTGC 268
 QY 305 GlnProGluLeuSerGluThrSerGlnThrLeuProProLysProPheSerCysGlyArg 324
 DB 269 CAGCCGAGCTTTCAGAGACATCCAGACTCTCCCAAGCCCTTCTCATGTGAGCGG 328
 QY 325 SerGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArgLysThrProVal 344
 DB 329 AGTGAAAGGAGCATAAAGAAATCCCATTTGGAAGTACAGAAAGAAAGACTCTCGTT 388
 QY 345 LysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysHisSerProIle 364
 DB 389 AAAAAGCTGCTCTCTAATTTTCAAAAAGTCAAAAACAAACCTCTAAGCACTCTCCATT 448
 QY 365 LysGluGluProCysGlySerLeuSerGluThrValCysLysArgGluLeuArgSerGln 384
 DB 449 AAAAGAGAACCTGTGTCTCTTATCTGAACCTGTTGAAGCTGAATAGAGAGCAA 508
 QY 385 GlnThrProGluLysProArgSerSerValAspThrProProArgLeuSerThrProGln 404
 DB 509 GAAACCCCAAGAAAGCCCGGCTTTCAGTGAGACCCCAAGCAAGCTCTCCACCCCA 568
 QY 405 LysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuPro 424
 DB 569 AAGGAGCCAGCAGCACCATCCCAAGAGAAAGCTTCTCAAGTAGATAGAAATTTGCCG 628
 QY 425 TyrLeuSerThrThrGluMetTyrLeuCysArgTyrHisGlnProProSerProLeu 444
 DB 629 TACCTTTCACCAACAGAAATGATTTGTGTGTGACACACAGCTCTCCCATCAGCTTA 688
 QY 445 ProLeuArgGluSerSerProLysLysGluGluThrValAla 458
 DB 689 CCAATTACGGGAATCTCTCCAAAGAAAGAGAGACTGTAGCA 730
 RESULT 13
 CN355144 704 bp mRNA linear EST 16-MAY-2004
 LOCUS 170060532602228 GRN_ES Homo sapiens cDNA 5', mRNA sequence.
 CN355144
 VERSION CN355144.1 GI:47355078
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 704)
 Brandenberger, R., Wei, H., Zhang, S., Lei, S., Muraige, J., Fisk, G., J.,
 Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,
 Lebkowski, J. and Stanton, L.W.
 Transcriptome characterization elucidates signaling networks that
 control human ES cell growth and differentiation
 Nat. Biotechnol. 22 (6), 707-716 (2004)
 CONTACT: Brandenberger R
 Regenerative Medicine


```

Db      181 AGTGAAGAGGACACAAAAGAAAACCCCATTTGGAAATACAGAAAGAAAGACTCTGTT 240
Qy      345 LyeLysLeuAlaProGluPheSerLysValIleThrLysThrProLysHisSerProIle 364
Db      241 AAAAAGCTGGCTCTGTAATTTTCAAAAGTCAAAACAAAACCTCTAAGCACTCTCCATT 300
Qy      365 LyeGluGluProCysGlySerLeuSerGluThrValCysLysArgGluLeuArgSerGln 384
Db      301 AAAGAGAACCTGTGGTTCATATCAGAAACTGTTTAAACGTGAATGAGAGCCAA 360
Qy      385 GluThrProGluLysProArgSerSerValAlaPheThrProArgLeuSerThrProGln 404
Db      361 GAAACCCCAAAAAGCCCGGTCTTCAGTGATACCCCAACCAACTCTGACTCCCA 420
Qy      405 LyeGlyProSerThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuPro 424
Db      421 AAGGAGCCACGACCCCAAGAGAAAGACCTTCTCAAGTATGATGAAATTTCCG 480
Qy      425 TyrluSerThrThrgluMetTyrluCyArgTrpHisGlnProProProSerProLeu 444
Db      481 TACCTTCCACACAGAAATGTATTG-TGTGTTGGACACAGCCTCCCATCAGCTTA 539
Qy      445 ProLeuArgGlu-SerSerProLysGluGluThrValAlaArgCysLeuMetProse 464
Db      540 CCAATTACGGGAAATCTCTCCAAAGAGAGAGACTGTAGCAAGGTCTGATGCCATC 599
Qy      464 rSerValAlaGlyGluThrSerValLeuAlaValProSerTrpArgAspHisSerValGly 484
Db      600 AAGTGTTCAGAGAAACCTTCAGTCTTGCTGTTCTTCTTGGAGGACCACTCAGTAGA 659
Qy      484 uProLeuArgAspProAsnProSerAspLeuLeuGluAsnLeuAspSerValPheSe 504
Db      660 GCCTTAAGGAGCAACCAATCTTCAGACAT-TTGGAGAACCTGATGACAGTATTTTC 718
Qy      504 r-LysArgHisAlaLys-LeuGluLeuAsp--GluLysArgGlyLys-ArgTrpAsp-I 522
Db      719 AAAAGAGCAGTGAACAACTTGAGCTTAAACGAGAAACGAGGAAACCGATGCGATAC 778
Qy      522 leglnArgIleArgGlu--GlnArgIleLeuGlnArgLeuGlnLeuArgMetTyrlu 541
Db      779 TCAGAGAGATCAGGAAACACAGAAATTAACAGCATGCGACTCAGATGTCTCCA 838
Qy      541 yslYslGlyIlegln 545
Db      839 CCACCGGAATCCGC 852

RESULT 15
LOCUS   BU059063              727 bp      mRNA      linear      EST 26-AUG-2002
DEFINITION   UI-M-FRO-cal-o-07-0-UI.r1 NIH BMAP_FRO Mus musculus cDNA clone
IMAGE:6414174 5', mRNA sequence.
ACCESSION   BU059063
VERSION     BU059063.1  GI:2249352
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LINL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)

```

```

FEATURES             Seq primer: pyx-5.
                     Location/Qualifiers
source               1..727
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6414174"
                     /cisse_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5,18.5,19.5,20.5,21.5,22.5,23.5,24.5,25.5,26.5,27.5,28.5,29.5,30.5,31.5,32.5,33.5,34.5,35.5,36.5,37.5,38.5,39.5,40.5,41.5,42.5,43.5,44.5,45.5,46.5,47.5,48.5,49.5,50.5,51.5,52.5,53.5,54.5,55.5,56.5,57.5,58.5,59.5,60.5,61.5,62.5,63.5,64.5,65.5,66.5,67.5,68.5,69.5,70.5,71.5,72.5"
                     /lab_pos="DH10B (T1 phage resistant)"
                     /clone_id="NIH_BMAP_FRO"
                     /note="Organ: Brain; Vector: pyx-Asc; Site 1: EcoR I;
                     Site 2: Not I; The library was constructed according
                     to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pyx-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAAGACAG. This library was created for the University
                     of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the
                     Developing Mouse Nervous System, supported by National
                     Institute of Mental Health (NIMH), Heman Chin, Ph.D.,
                     program coordinator."

ORIGIN
Alignment Scores:
Pred. No.:          5,7e-61          Length:          727
Score:              1192.00          Matches:          228
Percent Similarity: 97.47%          Conservative:      3
Best Local Similarity: 96.20%        Mismatches:       4
Query Match:        36.81%           Indels:           2
DB:                  5               Gaps:             2

US-10-054-935-2 (1-614) x BU059063 (1-727)
Qy      1 MetThrMetArgSerAlaValPheLysAlaAlaAlaProAlaGlyGlyAsnProGlu 20
Db      15 ATGACCATGAGATCCCAAGTTCAGAGCGCGCGCCCTCCCGGCAACCCCGAG 74
Qy      21 GlnArgLeuAspTrpGluArgAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
Db      75 CAGCGACTGACATCAAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 134
Qy      41 AlaGluAlaHisPheLeuProArgHisArgLysLeuLysGluProGlyProProLeuAla 60
Db      135 GCGGAGCCCATTTCTCCCGCGCGCATCGTAAGTCAAGAGCGCGCGCGCGCGCGCGCG 194
Qy      61 SerSerGlnGlyLysSerProAlaProSerProAlaGlyCys--GlyGlyLysArg 79
Db      195 TCTTCCAGGGCGGAGCGCCCTCCCGCTCCAGCGCGCTGCGCGCGCGCGCGCGCGCG 254
Qy      80 GlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGln 99
Db      255 GGCTGTGTTACTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 314
Qy      100 ValProLeuProCysProProProAlaThrLysGlnAlaGlyIleGlyGlyProAla 119
Db      315 GTGCGCTGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 374
Qy      120 AlaAlaGlyAlaGlyCysSerProArgProLysTrpGlnAlaValLeuProIleGlnThr 139
Db      375 GCGAGCGCGCGCTGCTGCGACCCCGCGCGCAAGTATCAGCGCGGTGCGCCATTCAAGC 434
Qy      140 GlySerLeu--ValAlaAlaLysGluProThrProTrpAlaGlyAspLysGlyGly 158
Db      435 GGCTCTATCGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 494
Qy      159 AlaLysSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuPro 178

```

Db 495 GGCGCTCCCCCAGCTGCGCACCGGCTGGAACCGGGGAGACCCACCACTACCTTGCCC 554

Qy 179 GlyProProProLeuAlaProThrAlaGlyThrLeuAlaAsnSerGlnGlyArg 198
|||||
555 GGCCCAACCAACCCCTCGCGCCACCGCACTGCTGGGACCTGGCGGCGCAGTGAGGCGAGA 614
|||||

Qy 199 TrpLysSerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAlaSerSerGln 218
|||||
615 TGGAAAGATATAAGAAAGAGCCCTCTCGGNGTGGCGCGGCTCGGGAGCCTCCAGTCAG 674
|||||

Qy 219 AlaAlaCysLeuLysGlnLeuLeuLeuGlnLeuAspLeuIleGlnGln 235
|||||

Db 675 GCCGCTGCTGCANACAGATCTTCTGCTGCAATTGGAACCTCATCGAACAG 725
|||||

Search completed: March 25, 2005, 23:02:26
Job time : 5434 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 03:02:34 ; Search time 7487 Seconds
(without alignments)
3973.755 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 614

Sequence: 1 MTRSAVFKAAPAGGNPE.....RSRRLRIKQKQPHRTCRK 614

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 15

Total number of hits satisfying chosen parameters: 27

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xjh
-Q=cgnt2/US10054935/rnat_23032005_144913_26366/app_query.fasta.1.775
-DB=GenEmbl -QFWT=faetap -SUFWT=oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=100
-DOCALLIGN=200 -THR SCORE=quality -THR MIN=15 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054935 @CGN 1.1.4198 @rnat_23032005_144913_26366 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -KONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	255	41.7	194780	9	AC068669 Homo sapi
2	188	30.6	655	6	CQ724830 Sequence
3	188	30.6	1290	6	AR379777 Sequence
4	155	25.2	3425	6	CQ842718 Sequence

5	155	25.2	3425	9	AK125654 Homo sapi
6	152	24.8	1269	9	BC039449 Homo sapi
7	143	23.3	580	11	BV167952 sgmm6889
8	143	23.3	1092	11	BV178745 sgmm10234
9	101	16.4	2745	9	BSM807332 Homo sapi
10	96	15.6	2638	9	AK124185 Homo sapi
11	73	11.9	2933	10	BC058629 Mus muscu
12	70	11.4	440	6	BC067002 Sequence
13	70	11.4	2087	10	BC055715 Mus muscu
14	70	11.4	2095	10	BC043039 Mus muscu
15	54	8.8	106773	2	AC119462_3 of
16	54	8.8	183618	2	AC113356 Rattus no
17	54	8.8	196724	10	AL590963 Mouse DNA
18	44	7.2	615	6	AX401049 Sequence
19	42	6.8	230913	6	AC119484 Rattus no
20	26	4.2	904	6	BD017820 Novel gen
21	26	4.2	904	6	BD097758 Novel gen
22	21	3.4	208620	2	AC107643 Mus muscu
23	19	3.1	60	6	CQ552685 Sequence
24	18	2.9	169993	2	AC148580 Gaetextost
25	17	2.8	903	6	BD018136 Novel gen
26	17	2.8	903	6	BD098074 Novel gen
27	16	2.6	162385	5	AL922903 Zebrafish

ALIGNMENTS

RESULT 1
AC068669
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 194780)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-749116
Unpublished
2 (bases 1 to 194780)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukanger,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardina,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Hatford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Melchior,J., Meneus,L., Mihova,T., Mirzadeh,C., Mieng,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testave,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS
DIRECT Submission
Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 194780)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukanger,B.,
Camatasto,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,

Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (14-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 194780)

REFERENCE
Birtten, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (17-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 17, 2003 this sequence version replaced gi:27734042.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

-----Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----Project Information
Center project name: LI0045
Center clone name: 749_I_16

FEATURES
source
1.194780
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/Chromosome="17"
/map="17"
/clone="RP11-749I16"
/clone_1lb="RPCT-11 Human Male BAC"
2.285
/rpt_family="L1MC3"
302.324
/rpt_family="(TTG)n"
529.637
/rpt_family="FLAM_A"
548.533
/note="<30 qual SNGP region"
repeat_region
complement(711..1007)
/rpt_family="AluSx"
complement(1038..1179)
/rpt_family="MER15"
2169.2473
repeat_region

/rpt_family="AluSp"
3642.3946
/rpt_family="AluY"
complement(4181..4391)
/rpt_family="AluJo"
4863.4901
/rpt_family="(GGGA)n"
5391.5716
/rpt_family="AluSx"
6700.6739
/rpt_family="(CA)n"
complement(7332..7491)
/rpt_family="MIR3"
7684.7805
/rpt_family="(CGGG)n"
7700.7856
/note="single clone coverage"
7788.7802
/note="<30 qual single clone coverage"
7825.7845
/note="<30 qual single clone coverage"
7931.7951
/rpt_family="GC_rich"
7980.8009
/rpt_family="GC_rich"
8388.8442
/rpt_family="GC_rich"
9670.9780
/rpt_family="L2"
10423.10557
/rpt_family="CT_rich"
11371.11675
/rpt_family="AluSx"
11724.11913
/rpt_family="MIR"
12708.13030
/rpt_family="AluJb"
13109.13189
/rpt_family="MIR3"
14044.14165
/rpt_family="CT_rich"
14184.14282
/rpt_family="CT_rich"
15860.15976
/rpt_family="CT_rich"
15894.16062
/rpt_family="C-rich"
17522.17592
/rpt_family="(CAGA)n"
18316.18420
/rpt_family="MIR"
21106.21188
/rpt_family="MIR"
complement(22670..22950)
/rpt_family="AluSx"
23005.23172
/rpt_family="MIR"
23173.23481
/rpt_family="AluJb"
complement(23420..23431)
/note="<30 qual SNGP region"
23482.23525
/rpt_family="MIR"
23629.23697
/rpt_family="MIRSA"
complement(23698..24000)
/rpt_family="AluSc"
24001.24113
/rpt_family="MERSA"
complement(24215..24483)
/rpt_family="AluSx"
complement(24529..24534)
/note="<30 qual SNGP region"

```
unsure      complement (24541..24551)
unsure      /note="<30 qual SINGL region"
unsure      complement (24561..24569)
unsure      /note="<30 qual SINGL region"
repeat_region 24586..24632
repeat_region /rpt_family="AluX"
repeat_region complement (24644..24770)
repeat_region /rpt_family="MIR"
repeat_region 24821..25072
repeat_region /rpt_family="AluSX"

Alignment Scores:
Pred. No.: 1,72e-236 Length: 194780
Score: 256.00 Matches: 256
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.69% Indels: 0
DB: 9 Gaps: 0

US-10-054-935-2 (1-614) x AC068669 (1-194780)

QY 1 MetTherMetArgSerAlaValPheLyAlaAlaAlaProAlaGlyYAsnProGlu 20
Db 67584 ATGACCATGAGATCCGGGTTCAGAGCCGCCGCCCTCCGCCGCAATCTGAG 67643

QY 21 G1AArgLeuAspTyrGuaArgAlaAlaAlaLeuGlyYProGluAspGluProGluAla 40
Db 67644 CAGCGACTGACATACGAGCGGGCTGCGCGCTGGCGGGCCCGAGACGAGCTGGGGCG 67703

QY 41 AlAG1uAlaHisPheLeuProArgHisArgLyLeuYsgLupProGlyProProLeuAla 60
Db 67704 GCCGAGCCACATTCCTCCCGGCCGACGTAAAGTCAAGAGCGGGGCCCGCTGGCC 67763

QY 61 SerSerGlnGlyYSerProAlaProSerProAlaGlyYCySgLyGlyYValArgGly 80
Db 67764 TCCTCCAGGCGGAGCCCGCCCTTCGCCCGCGCTGCGCGCAAGGCGCGGCGC 67823

QY 81 LeuLeuLeuProAlaGlyAlaAlaProGlyYngLingLupGlyYSerVal 100
Db 67824 TTGTACTCCCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 67883

QY 101 ProLeuProCySProProProAlaThrLySglnAlaGlyY1egLyGlyYProAlaAla 120
Db 67884 CCTTGGCCCTGTCGCCCGCGGCCGACCAAGCAACCGGCATTTGGGGGGAGCTGGCGCA 67943

QY 121 AlAG1YAlaGlyYCySProProArgProLySglnAlaValLeuPro11egLThrGly 140
Db 67944 GCCGAGCCGCGCTGCAAGCCCGCGCCCAAGTACAGCGGTGCTGCCATTCAAGCGGCG 68003

QY 141 SerLeuValAlaAlaAlaYsgLupProThrProThrProAlaGlyYAspLySgLyYAlaAla 160
Db 68004 TCTTCCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 68063

QY 161 SerProAlaAlaThrAlaSerAspProAlaGlyYProProProLeuProLeuProGlyPro 180
Db 68064 TCCCGCCCTGCAACCGCTCGGAGCCCGGGGAGCCCACTAATCTTCCCGGGCGG 68123

QY 181 ProProLeuAlaProThrAlaThrAlaGlyYThrLeuAlaAlaSerGlyYArgTyrLys 200
Db 68124 CCACCCCTCGGCGCCACCGCGCACCGCGGACCTGGCGGCGCACGAGGAGAGATGAGAG 68183

QY 201 SerMetArgLySAspProLeuGlyYGlyYGlyYGlyYValAspSerGlnAlaAla 220
Db 68184 AGATATGAGAGAGACCTCTCTGGGGGGTGGTGGCGGCTCGGAGGCTCCAGTCAAGCGCGC 68243

QY 221 CysLeuYsgLIn1eLeuLeuGlnLeuAspLeu11egLingLingLingLingLIneu 240
Db 68244 TGCTTCAAAACAGATCTTCTGCTCATTTGACCTCATCGAACAGCAGCAGCAGCAGCTG 68303

QY 241 GlnAlaYsgLupLySgLn11egLingLupLeuYsgLupArgAspThr 256
Db 68304 CAGGCAAGGAGAGAGATCGAGAGCTGAAGTCAAGAGAGAGAGAGAG 68351
```

```
RESULT 2
CQ724830 655 bp DNA linear PAT 03-FEB-2004
LOCUS Sequence 10764 from Patent WO02068579.
DEFINITION CQ724830
ACCESSION CQ724830
VERSION CQ724830.1 GI:42285687
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
AUTHORS Kils, such as nucleic acid arrays, comprising a majority of
TITLE humenexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 10764 06-SEP-2002;
FEATURES
location/Qualifiers
source 1..655
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
US-10-054-935-2 (1-614) x CQ724830 (1-655)

Alignment Scores:
Pred. No.: 1,27e-172 Length: 655
Score: 188.00 Matches: 188
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 30.62% Indels: 0
DB: 6 Gaps: 0

QY 332 LysSerProheGlySerThrGuaArgLySgLnProValYsLysLeuAlaProGluPhe 351
Db 2 AATGCCCATTTGGAGTACAGAAAGAAAGACTCTGTTAAAAAGCTGCTCGAATTT 61

QY 352 SerLySValYsThrLySThrProLySHisSerPro11eYsgLupProCySgLySer 371
Db 62 TCAAAAGTCAAAACAAAAACCTCTTAAGCACTCTCTTAAGAGGAACTGTGGTTCC 121

QY 372 LeuSerGlnThrValCySlySArgGlnLeuArgSerGlnGlnThrProGluYsgLupProArg 391
Db 122 TTATCTGAACCTGTTTGTAACTGAATTAAGAGCCAAAGAACCCCAAGAACCCCGG 181

QY 392 SerSerValAspThrProProArgLeuSerThrProGlnYsgLyProSerThrHisPro 411
Db 182 TCTTCAAGTGAACCCACCAAGAGCTCCACTCCCAAAAGGAGCCAGACCAATCCC 241

QY 412 LysGlnLysAlaPheSerSerGln11egLupLeuProTyrLeuSerThrThrGlnMet 431
Db 242 AAGGAGAAACCTTCTCAAGTGAATGAAGATTGGCTGTAACCTTCCACACAGAAATG 301

QY 432 TyrLeuCySArgTyrPheGlnProProProSerProLeuProLeuArgLySgLnSerPro 451
Db 302 TATTGTGTGTGGCCAGCAGCTCCCAATCAACCTTACCATTAACGGAATCTCTCCA 361

QY 452 LysLysGlnGlnThrValAlaArgCySLeuMetProSerSerValAlaGlyGlnThrSer 471
Db 362 AAGAGAGAGAGACTGACAGAGTGTGATGCAATGATGTGACAGAGAGAACTTCA 421

QY 472 ValLeuAlaValProSerTyrPArgAspHisSerValGlnProLeuArgAspProAspPro 491
Db 422 GTCTTGCTGTCTCTTCTTGAAGGAGCACTCAAGTAAAGCTTAAAGGAGCCCAAACTCT 481

QY 492 SerAspLeuGlnGlnLeuLeuAspAspSerValPheSerLySArgHisAlaYsLeuGln 511
Db 482 TCAAGCTTTTGGAGAACTGAGATGACAGTGTGTTTCAAGGCGGAGTCAAAATCTGAG 541

QY 512 LeuAspGlnYsArgArgLySArg 519
```

Db	542	CTGATGAGAGAGAGAGAAAGG	565
RESULT 3			
LOCUS	AR379777	1290 bp	DNA
DEFINITION	Sequence 322 from patent US 6607879.		linear
ACCESSION	AR379777		PAT 18-DEC-2003
VERSION	AR379777.1	GI:40087411	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 1290)		
TITLE	Cocks,B.G., Stuart,S.G. and Seilhamer,J.J.		
JOURNAL	Compositions for the detection of blood cell and immunological		
FEATURES	response gene expression		
Source	Patent: US 6607879-A 322 19-AUG-2003;		
	Location/Qualifiers		
	1..1290		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ALIGNMENT SCORES:			
Pred. No.:	2.35e-172	Length:	1290
Score:	188..00	Matches:	201
Percent Similarity:	99.50%	Conservative:	0
Best Local Similarity:	99.50%	Mismatches:	0
Query Match:	30.62%	Indels:	1
DB:	6	Gaps:	0
US-10-054-935-2 (1-614) x AR379777 (1-1290)			
Qy	258	Leu1aAArg1Ieg1uAArgMetG1uAArgMetG1uAVal1yAlaAspAnG1uBys	277
Db	3	CTTGCTCGAATTGAACGTATGAGAAAGCGGATGCAGCTGTAAAGAGATACAGAGAA	62
Qy	278	G1uAArgH1s1yS1euphEng1y1YrG1uThrG1uAArg1uG1uThrG1uAArgSer	297
Db	63	GAAAGGCAAGACTGTTTCAAGGCGTATGAACCTGAAGAGAGAGAGAAACAGACTATCT	122
Qy	298	G1u1yS1Ie1yS1e1uG1uCySg1nPrG1uAArgSerG1uThrSerG1nThr1eudPro	317
Db	123	GAGAAATTAATACTGAGTGCACGCGGAGCTTCCGAGACATCCAGACTCTGCTCC	182
Qy	318	LysProPheSerCysG1yAArgSer-G1y1yG1yH1s1yAArg1ySerProPheG1yS	337
Db	183	AAGCCCTTCTATGGGCGGAGTGGGAAAGGAGACATAAAGAAATCCCATTTGGAG	242
Qy	337	rThrG1uAArg1yThrProVal1yS1e1uA1aProG1uPheSer1yS1e1yS1e1yTh1y	357
Db	243	TACAGAAAGAAAGACTCCTGTTAAAGAAAGCTGCTCTCTGAATTTTCAAAAGTCAAAACAA	302
Qy	357	sThrPro1yS1h1sSerPro1Ie1ySg1uG1uProCySg1ySer1e1uSerG1uThrValCy	377
Db	303	AACTCCTAAGACCTCTCTATTAAGAGAAACCTGTGGTCTTATCTGAACCTGTTG	362
Qy	377	S1yAArg1u1e1uA1ySerG1nG1uThrProG1u1ySProAArgSerValAspThrPr	397
Db	363	TAAACGTAAATTGAGAGACCAAGAAACCCAGAAAAAGCCCCGCTTCAAGTGAACACCC	422
Qy	397	oProAArg1e1uSerThrProG1u1ySg1yProSerThrH1s1yS1e1ySg1u1yA1aPhSe	417
Db	423	ACCAAGACTCTTCACCTCCCAAAAGAGACCCAGACCCATCCCAAGAGAGAAAGCCCTTCC	482
Qy	417	rSerG1u1IeG1uA1e1uPro1y1e1uSerThrThrG1uMet1y1e1uCyAArgTrpH1	437
Db	483	AAGTAGAGTAGAAGTTTGCGCTACTTTCCACCAAGAAAGTATTTGTGTGCTTGGA	542
Qy	437	EG1nProProProSerPro1e1uPro1e1uAArg1uSerPro1yS1ySg1uG1uThrVa	457
Db	543	CCAGGCTCCCATCAACCGTTTACATTTCGAGGAATCTCTCCAAAGAGAGAGAGACTGT	602

Oy	457	lAla 458 603 AGCA 606
RESULT 4		
LOCUS	CQ842718	3425 bp DNA linear PAT 02-AUG-2004
DEFINITION	Sequence 1365 from Patent EP1440981.	
VERSION	CQ842718	
KEYWORDS	CQ842718..1 GI:50894505	
SOURCE	.	
ORGANISM	Homo sapiens (human)	
REFERENCE	Homo sapiens	
AUTHORS	BukariVola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1	
TITLE	Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J., Isono,Y., Nagai,K. and Irie,R.	
JOURNAL	Full-length human cdna Patent: EP 1440981-A 1365 28-JUN-2004; Research Association for Biotechnology (JRP)	
FEATURES	Location/Qualifiers	
source	1..3425 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"	
ORIGIN		
Alignment Scores:		
Pred. No.:	6.93e-140	Length: 3425
Score:	155.00	Matches: 155
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	25.24%	Indels: 0
DB:	6 Gaps:	0
US-10-054-935-2 (1-614) x CQ842718 (1-3425)		
Oy	460	CysLeuMetProSerSerValAlaGlyGluThrSerValIeuAlaValProSerTrpArg 479
Db	540	TGTCTGATGCCCATCAGTGTGGCGAGGAACCTTCAGTCTTGCTGTTCCCTTTGGAGG 599
Oy	480	AspHisSerValGluProlLeuArgAspProAsnProSerAspLeuGluAsnLeuAsp 499
Db	600	GACCACCTCAGTAGAGCCTCTAAGGGAGACCCAATCTTCAGACCTTTGGAGAACAACCTGCAT 659
Oy	500	AspSerValPheSerLysARGHISalaLysLeuGluLeuAspGluLysARGLysArg 519
Db	660	GACAGTGTGTTTTGAAAGCGGACGACAAACTTGAAGCTGGATGAGAGAGAGAAAGAAAGA 719
Oy	520	TrpAspIleGlnArgIleArgGluGlnArgIleLeuGlnArgLeuGlnLeuArgMetLys 539
Db	720	TGGGATATTTCAGAGATTCAGGGAACAAGATTTTACAGCGACTGCAGCTCAGATGAT 779
Oy	540	LysLysLysGlyIleGlnGluSerGluProlGluValThrSerPhePheProlGluProAsp 559
Db	780	AAAAAGAAAGAAATTCCAGGAATCGACGCTGAGGTTTACTCATTTTCCCCTGAGCCAGAT 839
Oy	560	AspValGluSerIleuMetIleThrProPheLeuProValAlaIlePheGlyArgProLeu 579
Db	840	GATGTGTGAATTTGATGATTAACCCCTCTTCTGCTGTGTAGGATTTGGACGACCATTA 899
Oy	580	ProLysLeuThrProGlnAsnPheGluLeuProTrpLeuAspGluArgSerArgCySarArg 599
Db	900	CCAAAATTAACTCCACAGAAATTTTGAGCTAACCTCGTTGGATGAGCGTAGCCGATGCAGA 959
Oy	600	LeuGluIleGlnLysLysGlnThrProHisArgThrCySarGlyLys 614
Db	960	TTGGAGATCCAGAAGAACAAACACTCACCGGACGTGTAGAGAA 1004
RESULT 5		
LOCUS	AK125654	3425 bp mRNA linear PRI 09-SEP-2003

DEFINITION	Homio sapiens cDNA FLJ43666 fls, clone SYN0V4007012.
ACCESSION	AKI25654
VERSION	AKI25654.1 GI:34531819
KEYWORDS	oligo capping, fls (full insert sequence).
SOURCE	Homio sapiens (human)
ORGANISM	Homio sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 Ota, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Ito, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kituchi, H., Kanda, K., Magatsuna, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Maehuo, Y., Negai, K. and Isogai, T.
TITLE	NEO human cDNA sequencing project
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 3425)
AUTHORS	Isogai, T. and Yamamoto, J.
TITLE	Direct Submission
JOURNAL	Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail:genominfo@hri.co.jp, Tel:81-438-52-3375, Fax:81-438-52-3986)
COMMENT	NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3'-end one pass sequencing: RAB HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
FEATURES	Location/Qualifiers
SOURCE	1..3425
	/organism="Homio sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="SYNOV4007012"
	/cissue_type="synovial membrane tissue from rheumatoid arthritis"
	/clone_lib="SYNOV4"
	/note="Cloning vector: pME18SFL3"
ORIGIN	
Alignment Scores:	
Pred. No.:	6.93e-140
Score:	155.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	25.24%
DB:	9
Gaps:	0
US-10-054-935-2 (1-614) x AKI25654 (1-3425)	
QY	460 CysleuWetProSerSeValAlaGluThrSerValLeuAlaValProSerTrypArg 479
Db	540 TGTCTGAGGCATCAAGTGTTCAGAGAAACTTCACTTGTGGCTTCCTTGAGAG 599
QY	480 AspHisSerValGluProLeuArgAspProAsnProSerAspLeuLeuGluAsnLeuAsp 499
Db	600 GACCACTAGTAGACCTCTTAAGGGACCCAAATCTTCAAGACTTTTGGAGAACTGGAT 659
QY	500 AspSerValPheSerLeuArgHisAlaLeuLeuGluLeuAspGluLysArgArgLysArg 519
Db	660 GACAGTGTGTTTTCGAAACGGCATCCAAACTCGAGCTGGATGAGAAAGAGAAAGCA 719
QY	520 TrpAspIleGlnArgIleArgGluGlnArgIleLeuGlnArgLeuGlnLeuArgMetTyr 539
Db	720 TGGGATATTTCAGAGCATCAGGGAAACAAAGAAATTTTACAGCGCACTCAGATGTAT 779
QY	540 LysLeuLysArgIleGlnGluSerGluProGluValThrSerPhePheProGluProAsp 559
Db	780 AAAAAGAAAGAAATTCAGAAATCTAAGCTGAGGTACTCAATTTTCTCTGAGCAAGT 839

Qy	Db	RESULT 6	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK	COMMENT	
Qy	560	Asp	1	Asp	1	Asp	1	Asp	1	Asp	1	Asp	1	Asp	1	Asp	1	Asp	1	Asp	1	Asp
Qy	840	GATCTTGAAGATTGATGATTTACCCCTTCTTCCCTGTTGATGATTTGACGACCATTA	899																			
Qy	580	Prolys	1	Prolys	1	Prolys	1	Prolys	1	Prolys	1	Prolys	1	Prolys	1	Prolys	1	Prolys	1	Prolys	1	Prolys
Qy	900	CCAAATTTAACTCCACAGAAATTTTGACCTGCTGGTTGATGATGCGTACGATGCGA	959																			
Qy	600	LeuGlu	1	LeuGlu	1	LeuGlu	1	LeuGlu	1	LeuGlu	1	LeuGlu	1	LeuGlu	1	LeuGlu	1	LeuGlu	1	LeuGlu	1	LeuGlu
Qy	960	TTGGAGATCCAGAGAGCAAAACCTCCACCGACGTGTAGGAAA	1004																			
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp	1269	bp
Qy	1269	bp	1269																			

RESULT 8
BV178745 1092 bp DNA linear STS 10-JUN-2004
LOCUS Bgmm102348 Human DNA (Sequenciom) Homo sapiens STS genomic, sequence
DEFINITION tagged site.
ACCESSION BV178745
VERSION BV178745.1 GI:48015232
KEYWORDS STS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1092)
AUTHORS Nelson,R.M., Martenillo,G., Kammerer,S., Hoyal,C.R., Shi,M.M.,
Cantor,C.R., and Braun,A.
TITLE Large-Scale Validation of Single Nucleotide Polymorphisms in Gene
Regions
JOURNAL Genome Res. (2004) In press
COMMENT Contact: Andreas Braun
Pharmaceuticals division
Sequenciom, Inc.
3595 John Hopkins Court, San Diego, CA 92121, USA
Tel: 18582029018
Fax: 18582029020
Email: abraun@sequenciom.com
Primer A: No primer sequence submitted
Primer B: No primer sequence submitted
STS size: 1092.

FEATURES
Source
1. .1092
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_id="Human DNA (Sequenciom)"
<1. .>1092

ORIGIN
STS
Alignment Scores:
Pred. No.: 1.14e-128 Length: 1092
Score: 143.00 Matches: 173
Percent Similarity: 98.86% Conservative: 0
Best Local Similarity: 98.86% Mismatches: 1
Query Match: 23.29% Indels: 2
Gaps: 0

US-10-054-935-2 (1-614) x BV178745 (1-1092)

QY 11 AAlAAlAAlAProAlAGlYgYAsnProGlUGlnArgrLeuAspTYrGIuArGAlAAlA 30
Db 2 GCCCGGCCCCCTGCGCGGCAATCTGAGACGACGACTGACGACGAGCGGGCTGCGGCG 61

QY 31 LeuGlYgYProGlUAspGluProGlYAlAAlAGlUAlAhIsPheLeuProArGrHIsArg 50
Db 62 CTGGGGCGGGCCCGAGAGAGAGAGCTGGGGCGGGCCGAGAGCCCACTTCTCCCGGACACGT 121

QY 51 LyeLeuLysGluProGlYProProLeuAlAAsrSerGlnGlyGlySerProAlAProSer 70
Db 122 AAGCTCAAGAGAGCGGGGCGGCGCTGCTCTCCCAAGGGCGGAGAGCCCGGCGCTTCC 181

QY 71 ProAlAGlYCySgYgYLYsGlyArGrGlyLeuLeuLeuProAlAGlYAlAAlAProGlY 90
Db 182 CCGGCGGGCTGCGGCGGAGAGGCGGGGCTTGTACTCCCGGCGGGGCGGCGGCGGCGG 241

QY 91 GlnGlnGlnGlnUserTTPGlyGlySerValProLeuProCySProProProAlAThrLys 110
Db 242 CAGCAGAGAGAGAGCTGGGGGCGGTTCGCTGCTTGCCTGCGGCGGCGGCGGCGGCGG 301

QY 111 GlnAlAGlYlLeGlyGlyGluProAlAAlAAlAGlYAlAGlYCySsrProArGrProLys 130
Db 302 CAAGCCCGGCAATGGGGGAGAGCTGCGGAGCGGCGGAGCGGCGGCGGCGGCGGCGG 361

QY 131 TyrGlnAlAValLeuProIleGlnThrGlySerLeuValAlAAlAAlAGlYGluProThr 150

Db 362 TATCAGCGCGGTGCTGCCATTCAGACGGGCTCTCTGCTGGCGGCGGCAAGAGCCTAACG 421

QY 151 ProTTPAlAGlY-AspLYsGlyGlyAlAAlAAsrProAlAAlAThrAlAAsrApproAl 170
Db 422 CCTGGGCGTGA-GAGCAAGGAGGAGGCGGCGGCTCCCGGCTGCCAGCGGCTGGAGCCGGC 480

QY 170 agLYProProProLeuProLeuProGlYProProProLeuAlA 184
Db 481 GGGAGCCCGCACCACTACTGCTGCGGGCGGCGGCAACCCCTCGCG 523

RESULT 9
HSM807332 2745 bp mRNA linear PRI 30-AUG-2003
LOCUS HSM807332
DEFINITION Homo sapiens mRNA; cDNA DKFZp686P24239 (from clone DKFZp686P24239).
ACCESSION BX647188
VERSION BX647188.1 GI:34366216
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2745)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Oanger,A.,
Fobo,G., Han,M. and Wiemann,S.
TITLE The German Human cDNA Consortium
JOURNAL Direct Submission
Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuerberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by ACOMA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp686P24239) is available at the RZPD in Berlin.
Please contact the RZPD: Resourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.

FEATURES
Source
1. .2745
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686P24239"
/tissue_type="human small intestine"
/clone_id="686 (synonym: hlc03). Vector pSport1_Sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
2707. .2712
2726

ORIGIN
polyA signal
polyA_site
2726

Alignment Scores:
Pred. No.: 1.8e-87 Length: 2745
Score: 101.00 Matches: 101
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.45% Indels: 0
Gaps: 0

US-10-054-935-2 (1-614) x HSM807332 (1-2745)

QY 514 GlnLYeArGrArGlySArGTTPAspIleGlnArGllEArGlnGlnArGllELeuGlnArg 533
Db 3 GAGAAAGAAAGAAAGATGGATATTCAGAGATCGAGAAACAAATTTTACACGGA 62

QY 534 LeuGlnLeuArGrMetTYrLYsLYsLYsGlylLeGlnGlnUserGluProGluValThrSer 553
Db 63 CTCGAGCTCAGATGATTAATAAAGAAAGAAATTCAGGAATCTGAGCTGAGTTACTCA 122

QY 554 PhePheProGluProAspAspValGluSerLeuMetIleThrProPheLeuProValVal 573
Db 123 TTTTTCCTGAGCAGATGATGATTGAAGTTGATGATTAACCCCTTCTGCTGTTGTA 182

QY	574	AlphagelgAaPProlePrLyseuThrProGlnAnpPegUlePrTripleusp	593
Db	183	GCATTGTGACACCACTTTCACAAATTACTCACAGAAATTTTGACTTACCTCGTTGGAT	242
QY	594	GlnArgSerArgCyArgLeuGluIleGlnIlySylsGlnThrProHlaArgThrCysArg	613
Db	243	GAGCGTGAAGCCGATGCAGATTGGAAGATCCAGAAAGAACCAACCTTCACCGGACGTTAGG	302
QY	614	Lys 614	
Db	303	AAA 305	
RESULT 10			
AK124185		2638 bp	mRNA linear PRI 30-JAN-2004
LOCUS			
DEFINITION			
AK124185			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Tanase,T., Imose,N., Takeuchi,K.,			
Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,			
Sekine,M., Oabayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,			
Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,			
Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagaetsuna,M.,			
Shitatori,A., Sudo,H., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,			
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Fuyuta,T.,			
Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsure,N., Sato,K.,			
Tanikawa,M., Yamazaki,M., Niinomiya,K., Ishbaeni,T., Yamashita,H.,			
Murakawa,K., Fujimori,K., Tani,H., Kimata,M., Watanabe,M.,			
Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,			
Yosida,M., Hottuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,			
Hara,H., Tanase,T., Nomura,Y., Togiyama,S., Komai,F., Hara,R.,			
Tateuchi,K., Arita,M., Imose,N., Mutsashino,K., Yunki,H., Oshima,A.,			
Saeaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,			
Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,			
Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,			
Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,			
Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,			
Kunagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,			
Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,			
Ozaki,K., Hiraio,M., Ohmori,Y., Kawabata,A., Hiki,T., Kobatake,N.,			
Inagaki,H., Ikema,Y., Okamoto,S., Ohtani,R., Kawakami,T.,			
Noguchi,S., Iton,T., Shigeta,K., Senba,T., Matsunura,K.,			
Nakajima,Y., Mizuno,T., Morinaga,M., Saeaki,M., Togashi,T.,			
Oyama,M., Hara,H., Watanabe,M., Komatsu,T., Mitsuhashi-Sugano,J.,			
Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,			
Nagase,T., Nomura,N., Kikuchi,H., Masuno,Y., Yamashita,R.,			
Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isegai,T., and Sugano,S.			
Complete sequencing and characterization of 21,243 full-length			
human cDNAs			
JOURNAL			
PIUMED			
REFERENCE			
AUTHORS			
TITLE			
14702039			
2			
Oeshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,			
Arita,M., Mutsashino,K., Yunki,H., Hara,H., Sugiyama,T., Irie,R.,			
Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,			
Isono,Y., Kawai,H., Ito,Y., Saito,K., Nishikawa,T., Kimura,K.,			
Yamashita,H., Matsuno,K., Nakamura,Y., Sekine,M., Kikuchi,H.,			
Kanda,K., Wagaetsuna,M., Murakawa,K., Kanehori,K., Sugiyama,A.,			
Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuno,Y., Nagai,K.			
and Isegai,T.			
NEDO human cDNA sequencing project			
Unpublished			
3 (bases 1 to 2638)			
Isegai,T. and Yamamoto,J.			
Direct Submission			
Submitted (15-JUL-2003) Takao Isegai, FLJ Project (HRI Team) ; 2-6-7			

COMMENT					
	(E-mail: genomics@krisari.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)				
	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry for Biotechnology (RAB) / cDNA library Research Association for Biotechnology (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.				
FEATURES					
source					
	1..2638 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /collection="THYM2033053" /tissue_type="thymus" /clone_id="THYM2" /note="Cloning vector: pME18SPFL3"				
ALIGNMENT SCORES:					
Pred. No.:	1.26e-82	Length:	2638		
Score:	96.00	Matches:	96		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	15.64%	Indels:	0		
Df:	9	Gaps:	0		
US-10-054-935-2 (1-614) x AK124185 (1-2638)					
Qy	519	ArgTTTApSpIleGlnAgtlLeArvgUgInAArglIleUenGlnArGlenuLeuArgMet	538		
Db	599	AGATGGGATATTTCAGAGATCAGGGAACAAAGATTTCACGCACGCCAGCTAGATG	658		
Qy	539	TyTtLysLySLySGYllLegIngLuSeSerGluProGluValThrSerPhePheProGluPro	558		
Db	659	TATATAAGANAGAAATTCAGGAATCGAGCCTAGAGTTACTTATTTTTCCCTGAGCCA	718		
Qy	559	AsPaSpAyAlguSeSerLeuMeLtIethrProPhelEuProVaLaIalaphegLYarGPro	578		
Db	719	GATCATGTGTAAGATTGATGATTAACCCTCTTGCTGCTGTGTAGCAATTGACGACCA	778		
Qy	579	LeuProLySLeuThrProGlnAenPheGluLeuProTrpLeuAapGluVnGserArGys	598		
Db	779	TTTACCAAATTAATTCACAGAATTTTGAGCTACCCGTGTGATGAGCCGTAGCCGATGC	838		
Qy	599	ArgLeuGluIleGlnLySLySGlntThrProHisArgThrCySArGLys	614		
Df	839	AGATTGGAGATCCAGAGAGAACCAACCTCACCCGACGCTGTAGGAAA	886		
RESULT 11					
BC058629	BC058629 2933 bp mRNA linear ROD 30-JUN-2004				
LOCUS	MUS musculus RIKEN CDNA 4121d02D02 gene, mRNA (CDNA clone MGC:64809				
DEFINITION	IMAGE:6402215), complete cds.				
ACCESSION	BC058629				
VERSION	BC058629.1 GI:35193129				
KEYWORDS	MGC.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2933)				
AUTHORS	Straussberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Sherman,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Bietow,K.H., Schaefer,C.F., Bhac,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J.J., Heish,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stadton,M., Soares,M.B., Bonaldi,M.F., Casavant,T.L., Schneetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S., Caminci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Adamiro,R.D., Mullaly,S.J., Boesch,S.A., McEwan,P.J.,				

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.C., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butlerfield, Y.S., Krzyzanski, M.I., Skalska, U., Smallus, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 2933)
 Strausberg, R.
 Direct Submission
 Submitted (22-SEP-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk
 Email: cgaps-rc@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 CDNA Library Preparation: M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIMU)
 DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
 Thomas L. Casavant.
 Web site: http://genome.uiowa.edu
 Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
 Ronaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,
 Fishler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
 Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J.,
 Casavant, T., Soares, M.B.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LIML at: http://image.liml.gov
 Series: Plate: Row: Column: 0
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA g1: 21311987.
 Location/Qualifiers
 1. 2933
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="MGC:64809 IMAGE:6402215"
 /tissue_type="Mouse, brain 12.5 dpc"
 /clone_id="NIH_BMAP_FCO"
 /lab_host="DH10B"
 /note="Vector: pYX-ASG"
 1. 2933
 /gene="4121402D02R1k"
 /note="synonym: 2810017P12R1k"
 /db_xref="locusid:74026"
 /db_xref="MGI:1921276"
 63. 1913
 /gene="4121402D02R1k"
 /codon_start=1
 /product="RIKEN CDNA 4121402D02"
 /protein_id="AAHS6629.1"
 /db_xref="GI:35193130"
 /db_xref="MGI:1921276"
 /transluc="MTWRSVAFKAAAAPAGNPPORLDYERAAALGSPEDSGAAEAH
 FLPRRLKEGPPPLASGGSPSPAGCGGRGLLPAGAAPQOESWAGSVP
 LCPBPATKQAGIGEPVAAAGCSPKQYAVLPIDMGIVVAAAEPPWADKGS
 AAPPAATASDPAGPPPLPDPPLPLATATGTAAASGRKSRKRPPLGGGSGGS
 SOAACLOILLOLDLLEQQOQLOAKKEKELELSERDITLATEMERWQVKSD
 NEKERHKLQGEYBEERBEALSEKIKLERQPELCETSOALPSKPFSGRGKHKK
 TPENRERKTPVKLADPEPSKVTKTPHSGIKKEPCGSIETVCKEELSOETPER
 RSVYDTPRLSTPQKGSHPKREKAFSEMDLPLYLSTTEWYLCRMHOPPSPLPR
 SSPKKEETVACLMPSSVAGSTVLAVPSWMDHVEPLRDNPEDLLENDDSVFSKX

ORIGIN

Alignment Scores:

Pred. No.: 3 2e-60
 Score: 73.00
 Percent Similarity: 97.398
 Best Local Similarity: 97.398
 Query Match: 11.898
 DB: 10
 Gaps: 0

US-10-054-935-2 (1-614) x BC058629 (1-2933)

Qy 338 ThrGluArgLysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLys 357
 |||||
 Db 1080 ACAGAAAGAAAGACTCTTTAAAGCGCGCTCCGAAATTTTCAAAAGTCAAAACAAA 1139
 Qy 358 ThrProLysHisSerProLysGluGluProCysGlySerLeu-SerGluThrValCys 377
 |||||
 Db 1140 ACTCTTAACGACCTCTCCATTAAAGAGAACCTGTGTTCAT-ATCAGAAACTGTTTG 1198
 Qy 377 slySarGluLysLeuArgSerGluGluThrProGluLysProArgSerSerValAspThrPr 397
 |||||
 Db 1199 TAAACGTGAATTGAGAGCCAGAAACCCAGAAAGCCCGGTCTTCAGTGATACCC 1258
 Qy 397 oProArgLysSerThrProGluLysGlyProSerThrHisProLysGluLysAlaPhe 417
 |||||
 Db 1259 ACCAAGACTCTGACTCCCAAAAGGAGCCAGACCCACCCCAAGAGAAACCTTCTC 1318
 Qy 417 rSerGluLe-GluAspLeuProLysThrLysSerThrThriguMetYrLeuCysArgTrp 437
 |||||
 Db 1319 AAGTGAGAT-GGAAGATTGGCCGTACCTTCCACACAGAAAGTATTTGTGTGTTGGC 1377
 Qy 437 IsGluProProProSerProLeuArgLysSerSerProLysGluGluThrV 457
 |||||
 Db 1378 ACCAGCTCCCTCCATCACCGTTACATTACGGAATCTCTCCAAAGAGAGAGACTG 1437
 Qy 457 alAlaArgCysLeuMetProSerSerValAlaGlyLysThrSerValLeuAlaValPro 477
 |||||
 Db 1438 TACCAAGGTCTTGATGATCATCAAGTGTTCAGAGAAACTTACGCTTGCTTCTT 1497
 Qy 477 eTrpArgAspHisSerValGluProLeuArgAspProAsnProSerAspLeu-LeuGlu 496
 |||||
 Db 1498 CTGGAGGAGCACTCAGTACAGACCTCTAAGGAGCCCAATCTTCAGACAT-TTGGAG 1556
 Qy 497 AsnLeuAspAspSerValPheSerLysArgHisAlaLysLeuGluLeuAspGluLysArg 516
 |||||
 Db 1557 AACCTGATGACAGTGTATTTTCAAGAGGACATGCAAACTGGAGCTAGACGAGAGAG 1616
 Qy 517 ArgLysArgTrpAspLysLeuArgLysLeuArgLysLeuArgLysLeuArgLysLeu 536
 |||||
 Db 1617 AGAAACGATGATGATTTTCAGAGATCAGGAAACAAAGATTTTACGACATGACGCTC 1676
 Qy 537 ArgMetYrLysLysLysGlyLysGluLysGluSerGluProGluValThrSerPhePhePro 556
 |||||
 Db 1677 AGAATGATTAAGAAAGAAATTCAGGAATCTGAGACTGAGGTACTTATTTTCCCT 1736
 Qy 557 GluProAspAspValGluSerLeu 564
 |||||
 Db 1737 GAGCGATGATGTTGAAGTTTG 1760
 RESULT 12
 LOCUS CQ687002 440 bp DNA linear PAT 03-FEB-2004
 DEFINITION Sequence 31928 from Patent WO02070737.
 ACCESSION CQ687002
 VERSION CQ687002.1 GI:42217066
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
AUTHORS Liew, C.C., Marshall, W.E. and Zhang, H.
TITLES Compositions and methods relating to osteoarthritis
JOURNAL Patent: WO 02070737-A 31928 12-SEP-2002;
Chondrogene Inc. (CA)
FEATURES Location/Qualifiers
source 1..440
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 4,72e-58 Length: 440
Score: 70.00 Matches: 70
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.40% Indels: 0
DB: 6 Gaps: 0
US-10-054-935-2 (1-614) x CQ687002 (1-440)
QY 389 LysProArgSerSerValAspThrProProArgLeuSerThrProGlnYsGlyProSer 408
|||||
DB 7 AGCCCCGGCTTCACTGACACACCCACCAAGACTCTCCACTCCCAAGAGGACCCAGC 66
QY 409 ThrHisProIysGlnYsAlaPheSerSerGlnIleGlnAspleuProIYrIleuSerThr 428
|||||
DB 67 ACCCATCCCAAGAGAGAAAGCCCTTCTCAAGTAGATAGAAATTTGGCGTACCTTTCACAC 126
QY 429 ThrGluMetIYrIleuCysArgTrpHisGlnProProSerProIeuProIeuAArgGlu 448
|||||
DB 127 ACGAAGATGATTGTGTCGTGTGACACACCCCTCCCATCCGTTACCATTAACGGGAA 186
QY 449 SerSerProIysIysGlnGluThrValAla 458
|||||
DB 187 TCCTCTCCAAAGAGAGAGAGACTGTAGCA 216
RESULT 13
BC055715 2087 bp mRNA linear ROD 21-OCT-2003
LOCUS BC055715 2087 bp mRNA linear ROD 21-OCT-2003
DEFINITION IMAG:6414174), complete cds.
ACCESSION BC055715 GI:33585901
VERSION BC055715.1
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2087)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Bueltow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stepleten, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schaefer, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Muliyil, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gamaralle, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalski, U., Smalins, D.E.,
Scherich, A., Schen, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257

PUBMED 12477932
REFERENCE 2 (bases 1 to 2087)
AUTHORS Strausberg, R.
TITLES Direct Submission
JOURNAL Submitted (01-AUG-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) md@paxill.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAC Plate: 125 Row: m Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21311987.
FEATURES
source
Location/Qualifiers
1..2087
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:66782 IMAGE:6414174"
/tissue_type="Brain, pool of enriched mouse brain tissues
from 13.5, 14.5, 16.5 and 17.5 dpc"
/clone_1ib="NIH BMAP_FR0"
/lab_host="DH10B"
/note="Vector: pYX-ASC"
1..2087
/gene="4121402D02R1k"
/note="synonyms: 4930463F05R1k, 2810017F12R1k"
/db_xref="LOCUSID:74026"
/db_xref="MGI:1921276"
15..1406
/gene="4121402D02R1k"
/codon_start=1
/product="4121402D02R1k protein"
/protein_id="AAH55715.1"
/db_xref="GI:33585902"
/db_xref="LOCUSID:74026"
/db_xref="MGI:1921276"
/translation="MTMRSAVFKAAPAGNPNORLDYERAAALGGPEDESGAEEAH
FLPRHRKLKEPPPLASSGGSPSPGACGGKGLLPGAAAGQOEEGSGV
LPPRPATKQATIGEPVAGCSRPKYQAVLPOTGSIVAAAKEPTPNAGDGG
AAPPATADDPGPPPLPPPPAPATATATLAAISBRWKSIRKSPGGGSGSAS
SQALCKQLLLQLDLITEQOQOOLQKEIEISELSERDLARLERMRQVLKRD
NFKERKLTQYEPERERAEISEKIKERPELCETSGALPKSPSCSGSGNRRK
TFEGNERTKTPYKALAPERSKVTIKPKSPKIEKCGSISFTVCYRELRSGETPEPK
RSSVDPLPSTLQKGPSTHPREKAFSSMEDLPLVLTETMYLCRWHPPPSPPLURE
SSPKKEETVATASA"
ORIGIN
Alignment Scores:
Pred. No.: 1.94e-57 Length: 2087
Score: 70.00 Matches: 110
Percent Similarity: 98.21% Conservative: 0
Best Local Similarity: 98.21% Mismatches: 1
Query Match: 11.40% Indels: 2
DB: 10 Gaps: 0
US-10-054-935-2 (1-614) x BC055715 (1-2087)
QY 162 ProAlaIaThrAlaSerProAlaGlyProProProIeuProIeuProGlyProPro 181

```

Db      504 CCAAGCTCCACCGCTCGGAGCCCGCGGAGCCCGACACACTCTCGCCCGGCGCACCA 563
Qy      182 ProteulabProthralatrralaglyThleuAlalasergluglyargtTpylser 201
Db      564 CCCCCTCCGCGCCACCGCAGCTGCGGAGCCCTGCGCGCAGTGTGAGAGTGAAGAGT 623
Qy      202 Met-ArglyseerProleuglygllygllygllyserglYAlaserseglAlaacy 221
Db      624 AT-AAGGAAGAGCCCTCTCGGGGGTGGCGCGGCTCGGAGCCTCCAGTCAAGCCGCTG 682
Qy      221 sLeuylsGlnlleuleuleuInleuApleuileglunglnglnglnglnglng 241
Db      683 CCTCAACAGATCCTTCTGCTGCAATTGACCTCATGAAACAGACAGACAGCAGTTGCA 742
Qy      241 nAlalygllygllygllygllygllygllygllysergluArgasphrleuAlaag 261
Db      743 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 802
Qy      261 egluArgMetgluArgMetgluArgMetgluVallys 272
Db      803 TGAACGATATGGAAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 836
RESULT 14
BC043039 2095 bp mRNA linear ROD 21-OCT-2003
LOCUS Mus musculus RIKEN cDNA 4121402D02 gene, mRNA (cDNA clone MGC:57934
DEFINITION IMAGE:5701204), complete cds.
ACCESSION BC043039
VERSION BC043039.1 GI:27696670
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,M., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Sapich,T.E., Brownstein,M.J., Utsid,T.B., Toshiyuki,S.,
Schneitz,T.E., Brownstein,M.J., Utsid,T.B., Toshiyuki,S.,
Chenici,P., Prange,C., Raha,S.S., Loguclano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Boeak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,U.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.W., Gay,L.J., Hulyk,S.W.,
Vallalton,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J.J., Heltan,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whitling,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalhus,D.E.,
Schnierch,A., Schein,J.E., Jones,S.J., and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL MEDLINE
PUBMED 12477932
REFERENCE 2 (bases 1 to 2095)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.

```

```

Thomas L. Casavant.
Web site: http://genome.uiowa.edu/tom-casavant@uiowa.edu
Contact: bento-soares@uiowa.edu, tom-casavant@uiowa.edu
Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
Schneitz,T., Smith,C., Snir,E., Tack,D., Trouk,K., Walters,J.,
Casavant,T., Soares,M.B.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21311987.
location/Qualifiers
1. 2095
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:57934 IMAGE:5701204"
/tissue_type="Brain, mouse 15.5 dpc"
/clone_id="NIH_BMAP_EY0"
/lab_host="DH10B"
/lab="Vector: pYX-ASC"
/notes="Vector: pYX-ASC"
1. 2095
/gene="4121402D02Rik"
/notes="synonyms: 4930463F05Rik, 2810017F12Rik"
/db_xref="LOCUSID:74026"
/db_xref="MGI:1921276"
41. 1432
/gene="4121402D02Rik"
/codon_start=1
/product="4121402D02Rik protein"
/protein_id="AAH43039.1"
/db_xref="GI:27696671"
/db_xref="MGI:1921276"
/db_xref="MGI:1921276"
/translacion="MTMSAVFPAKAAAGNPNORLUDYERAAALGGPEDESGAAEH
FLPRHKLKRGEPPLASOGGSPSPACGGGKGGGLIPAAACGODESGGASV
LPSPPRKAGIGGSPVAGAGCSRPKQAVLPOTGIVVAAKEPFPMDKGG
AAPPAATVADPAPPPPLPLPGPPPLAPTATGTLASBGKWSIRKSPICGGGSGAS
SQAAPKQLLDLDLIDQOQDLQAKKEKIEBELKSBDRTLARIRMERMDLVKRD
NEKERHLQGYPEREERAELESEKIKLEKPELCTSGSLPSPKPSCGSGGKRRK
TPGNTERTKTPVKKLAPERSKVKTKTPKPSPIKEBPQGISETVCKRELRSGEPKP
RSVDPPPLSTSTOKPSTHPKRAKASSMEDLPYSTTMYLCRNHQPSPPLPLRE
SSEPKETVASKA"
ORIGIN
Alignment Scores:
Pred. No.: 1,95e-57 Length: 2095
Score: 70.00 Matches: 110
Percent Similarity: 98.21% Conservative: 0
Best Local Similarity: 98.21% Mismatches: 1
Query Match: 11.40% Indels: 2
DB: 10 Gaps: 0
US-10-054-935-2 (1-614) x BC043039 (1-2095)
Qy 162 ProAlaAlaThralaserAapProAlaaglyProProProleuProleuProgiProPro 181
Db 530 CCAAGTCCACCGCTCGGAGCCCGGCGGAGCCCGACACACTCTGCGCGGCGCACCA 589
Qy 182 ProteulabProthralatrralaglyThleuAlalasergluglyargtTpylser 201
Db 590 CCCCCTCCGCGCCACCGCAGCTGCGGAGCCCTGCGCGCAGTGTGAGAGTGAAGAGT 649
Qy 202 Met-ArglyseerProleuglygllygllygllygllyserglYAlaserseglAlaacy 221
Db 650 AT-AAGGAAGAGCCCTCTCGGGGGTGGCGCGGCTCGGAGCCTCCAGTCAAGCCGCTG 708
Qy 221 sLeuylsGlnlleuleuleuInleuApleuileglunglnglnglnglnglng 241
Db 709 CCTCAACAGATCCTTCTGCTGCAATTGAGCCTCATGAAACAGACAGACAGCAGTTGCA 768

```


constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6.

FEATURES
source
Location/Qualifiers
1..196724
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-395E10"
/clone_11b="RPC1-23"

ORIGIN

Alignment Scores:

Pred. No.:	4,34e-40	Length:	196724
Score:	54.00	Matches:	94
Percent Similarity:	97.92%	Conservative:	0
Best Local Similarity:	97.92%	Mismatches:	1
Query Match:	8.79%	Indels:	2
DB:	10	Gaps:	0

US-10-054-935-2 (1-614) x AL590963 (1-196724)

QY 162 ProAlaAlaThrAlaSerAspProAlaGlyProProProleuProleuProGlyProPro 181
DB 62257 CCAAGTGCACCGCTCGACCGCGGAGACCCACCACTTACTCTGCCGCGCACCA 62198
QY 182 ProleuAlaProThrAlaThrAlaGlyThrleuAlaAlaSerGluGlyArgTPLYSer 201
DB 62197 CCCCTGCGCGCCACCGCACTGCTGGGAGCCCTGCGCGCCAGTAGGGCAATGGAAGAT 62138
QY 202 Met-ArglySerProleuGlyGlyGlyGlySerGlyAlaSerSerGlnAlaAla 221
DB 62137 AT-AAGGAAGAGCCCTCTCGGAGGTGGCGCGCTCGGAGCCCTCCAGTCAGCCGCTG 62079
QY 221 sleuLYSGlnIleuLeuLeuGlnleuAspLeuIleGlnGlnGlnGlnGln 241
DB 62078 CCTCAACAGATCTTCTGCTGCATTGACCTCATCAACAGACGACGACGAGTTGCA 62019
QY 241 nalalySGluLYSGlnIleGlnIleuLeuLYSerGluArgASPThr 256
DB 62018 GGCCAAAGAGAGATAGAGAGGCGGAGAGTCCGAGAGAGATACG 61973

RESULT 18
AX401049 AX401049 615 bp DNA linear PAT 06-JUN-2002
LOCUS Sequence 725 from Patent WO0210453.
DEFINITION AX401049
ACCESSION AX401049
VERSION AX401049.1 GI:21337229
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS Wendick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and Blaschoff,M.R.
TITLE Molecular toxicology modeling
JOURNAL Patent: WO 0210453-A 725 07-FEB-2002;
Gene Logic, Inc. (US)

FEATURES
source
Location/Qualifiers
1..615
/organism="Rattus norvegicus"
/mol_type="unassigned DNA"
/db_xref="taxon:10116"
/note="EMBL/Genbank Accession No. A1102093"

ORIGIN

Alignment Scores:
Pred. No.: 1.21e-32 Length: 615
Score: 44.00 Matches: 125
Percent Similarity: 96.90% Conservative: 0

Best Local Similarity: 96.90% Mismatches: 2
Query Match: 7.17% Indels: 4
DB: 6 Gaps: 0

US-10-054-935-2 (1-614) x AX401049 (1-615)

QY 75 GLYGlyLYSGlyYArgGlyleuLeuProAlaGlyAlaAlaProGlyGlnGlnGln 94
DB 192 GGGGGCAAGGGCGGGGGTTGTACTCCGGCCGGGGGGCCCCCGGAGGAGGAG 251
QY 95 SerTPrGlyGlySerValProleuProCyProProProAlaThrIySGlnAlaGly 114
DB 252 AGCTGGGGGGGTTCCGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 311
QY 115 GLYGlyGlnProAla-AlaAlaGlyAlaGlyCySerProArgProLYTyrGlnAla 134
DB 312 GGGGGGAGACCGAGT-CGACAGCGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 370
QY 134 lleuProIleGlnIleuValAlaAlaLYSGlyGlnProThrProThrAlaGly 154
DB 371 GCTGCCATTCAAGCGGCTCTCTGCTGCGGGGCGCAAGAGAGCTTACGCTGGCTG 430
QY 154 YAspLYSGlyGlyAlaAlaSer-ProAlaAlaThrAlaSerAspProAlaGlyProPro 174
DB 431 GACCAAGGTTGGGGCGGC-TCCCCAGCTGCCACCGCTCGACCGGGGAGACCCCA 489
QY 174 roleuProleuProGlyProProProleuAlaProThrAlaThrAlaGlyThrleuAla 194
DB 490 CACTTACTTGGCCCGGGCGCCACCCCTCGCGCCGCCCACTGCTGGGAGACCTTGGCG 549
QY 194 laSerGlnIleuArgTPrLYSer 201
DB 550 CCAGTAGGCGCAGATGAAGAT 572

RESULT 19
AC119484/c 230913 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-38A24, *** SEQUENCING IN PROGRESS
DEFINITION
ACCESSION AC119484
VERSION AC119484.5 GI:25095516
KEYWORDS HTG: HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS Muzny,D.,Marie, Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benhamed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Casar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Hawlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M., Hollins,B., Howells,S., Huily,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Louieged,H., Lozado,R.J., Lu,X., Ma,J.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished
 2 (bases 1 to 230913)
 Direct Submission
 Submitted (27-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 230913)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23269735.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GVPB
 Center clone name: CH230-38A24
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990129
 Consensus quality: 218750 bases at least Q40
 Consensus quality: 222810 bases at least Q30
 Consensus quality: 224623 bases at least Q20
 Estimated insert size: 234093; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

FEATURES	source	Location/Qualifiers
* * * * *		* Runs of N, but the exact sizes of the gaps are unknown.
* * * * *		* This record will be updated with the finished sequence
* * * * *		* as soon as it is available and the accession number will
* * * * *		* be preserved.
* * * * *		* 1 173284: contig of 173284 bp in length
* * * * *		* 173285 173384: gap of unknown length
* * * * *		* 173385 173384: contig of 13940 bp in length
* * * * *		* 205325 205424: gap of unknown length
* * * * *		* 205425 207184: contig of 1760 bp in length
* * * * *		* 207185 207284: gap of unknown length
* * * * *		* 207285 210055: contig of 2771 bp in length
* * * * *		* 210056 210155: gap of unknown length
* * * * *		* 210156 212001: contig of 1846 bp in length
* * * * *		* 212002 212101: gap of unknown length
* * * * *		* 212102 230913: contig of 18812 bp in length.
FEATURES	source	Location/Qualifiers
		1..230913
		/organism="Rattus norvegicus"
		/mol_type="genomic DNA"
		/db_xref="taxon:10116"
		/clone="CH230-38A24"
		173385..174697
misc_feature		/note="wgs_contig"
misc_feature		175767..177669
misc_feature		/note="wgs_contig"
ORIGIN		
Alignment Scores:		
Pred. No.:	2.33e-28	Length: 230913
Score:	42.00	Matches: 42
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	6.84%	Indels: 0
DB:	2	Gaps: 0
US-10-054-935-2 (1-614) X AC119484 (1-230913)		
QY	519	ArgTTPApIIeGIInArgIIeArgGIuGIInArgIIeLeuGIInArgLeuGIInLeuArgMet 538
Db	211862	AGATGGGATATTTCAGAGCATCAGCGGACAAAGAAATTTTACACGCGACTGCAGGCTCAGAATG 211803
QY	539	TyrtYsIySlySgGIyIIeGIInGIsSerGIuPProGIuValThSerPhePhePProGIuPro 558
Db	211802	TATTAAGAAAGAAAGATTCAGAGATCTGAGCTGAGGTACTCATTTTCCCTGAGCCC 211743
QY	559	AspAp 560
Db	211742	GATGAC 211737
RESULT 20		
BD017820		
LOCUS	BD017820	904 bp DNA linear PAT 27-AUG-2002
DEFINITION	BD017820	Novel gene and novel gene fragment cloned in human neuroblastoma.
ACCESSION	BD017820	GI:22558996
VERSION	BD017820.1	GI:22558996
KEYWORDS	JP 2001245671-A/58.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
	1 (bases 1 to 904)	
REFERENCE	Nakagawara, A.	
AUTHORS	Novel gene and novel gene fragment cloned in human neuroblastoma	
TITLE	Patent: JP 2001245671-A 58 11-SEP-2001;	
JOURNAL	CHIBA PREF. HISAMITSU PHARMACEUTICAL CO INC	
COMMENT	OS Homo sapiens (human)	
	PN JP 2001245671-A/58	
	PD 11-SEP-2001	
	PF 07-MAR-2000 JP 2000159195	
	PI AKIRA NAKAGAWARA	
	PC C12N15/09, C12Q1/68, G01N33/53, G01N33/566//C12Q1/68, C12R1:91),	
	PC C12N15/00	
	CC Novel gene and novel gene fragment cloned in human CC	

neuroblastoma

FEATURES
source
Location/Qualifiers
1..904
/organism="Homo sapiens (human)"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 5,46e-15 Length: 904
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: 6 Gaps: 0

US-10-054-935-2 (1-614) x BD017820 (1-904)

QY 589 LeuProTipleuaspGluArgSerArgCysArgLeuGluIleGlnIlysgIntHrPro 608
|||||
Db 53 CTACCCGTGGTGGATGAGCGTACGCCATGCAGATTGGAGATCCGAGAGAACCAACACT 112
|||||

QY 609 HlaArgThrCysArgIlys 614
|||||
Db 113 CACCGAGCGTGTAGGAAA 130
|||||

RESULT 21
BD097758 904 bp DNA linear PAT 27-AUG-2002
LOCUS
DEFINITION Novel genes cloned in humanneuroblastoma and fragments thereof.
ACCESSION BD097758
VERSION BD097758.1 GI:22643332
KEYWORDS WO 0166719-A/58.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 904)
Nakagawara, A.
Novel genes cloned in humanneuroblastoma and fragments thereof
Patent: WO 0166719-A 58 13-SEP-2001.
CHIBA PREF. HISAITSU PHARMACEUTICAL CO INC, AKIRA NAKAGAWARA
OS Homo sapiens (human)
PN WO 0166719-A/58
PR 13-SEP-2001
PR 02-MAR-2001 WO 2001JP001629
PR 07-MAR-2000 JP 00P 159195
PI AKIRA NAKAGAWARA
PC C12N15/11, C12Q1/68, G01N33/53, G01N33/566
CC Novel genes cloned in humanneuroblastoma and fragments thereof
FH Key Location/Qualifiers
FT source 1..904
/organism="Homo sapiens (human)"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
Location/Qualifiers
1..904
/organism="Homo sapiens (human)"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 5,46e-15 Length: 904
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: 6 Gaps: 0

US-10-054-935-2 (1-614) x BD097758 (1-904)

QY 589 LeuProTipleuaspGluArgSerArgCysArgLeuGluIleGlnIlysgIntHrPro 608
|||||
Db 53 CTACCCGTGGTGGATGAGCGTACGCCATGCAGATTGGAGATCCGAGAGAACCAACACT 112
|||||

QY 609 HlaArgThrCysArgIlys 614
|||||
Db 113 CACCGAGCGTGTAGGAAA 130
|||||

RESULT 22
AC107643 208620 bp DNA linear HTG 22-MAR-2004
LOCUS
DEFINITION Mus musculus chromosome 14 clone RP23-54L14 map 14, *** SEQUENCING
IN PROGRESS ***, 7 ordered pieces.
AC107643
AC107643.4 GI:45599009
VERSION HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 208620)
Unpublished
2 (bases 1 to 208620)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campiano, S., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collimore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lacombe, K., Lamazares, R.,
Lander, T., Lehoczeky, J., Levine, R., Liu, G., Maclean, C.,
Macedonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrim, J., Menus, L., Mihova, T.,
Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, V., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 208620)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Atsachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collimore, A., Cook, A., Cooke, P., Dorris, L., Erickson, J., Faro, S.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lander, T., Levine, R.,
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C.,
Macedonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrim, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramaamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Scubs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAR-2004) Whitehead Institute/MIT Center for Genome

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 22, 2004 this sequence version replaced gi:29171454.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: U14249
 Center clone name: 54_L_14

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 *
 1 96362: contig of 96362 bp in length
 * 96363 96462: gap of 100 bp
 * 96463 99202: contig of 2740 bp in length
 * 99203 99302: gap of 100 bp
 * 99303 155557: contig of 56255 bp in length
 * 155558 155657: gap of 100 bp
 * 155658 165735: contig of 10078 bp in length
 * 165736 165835: gap of 100 bp
 * 165836 168011: contig of 2176 bp in length
 * 168012 173215: contig of 5104 bp in length
 * 173216 173315: gap of 100 bp
 * 173316 208620: contig of 35305 bp in length.
 Location/Qualifiers
 1. 208620
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="14"
 /map="14"
 /clone="RP23-54144"
 /clone_1id="RP23-54144"
 /clone_1id="RP23-54144" Female Mouse BAC"

FEATURES

Source

ORIGIN

Alignment Scores:

Pred. No.: 5.56e-08 Length: 208620
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conserved: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.42% Indels: 0
 Gaps: 0

US-10-054-935-2 (1-614) x AC107643 (1-208620)

QY 338 ThrGluArgLysThrProValLysLysLeuAlaProGluLysPheSerLysValLysThrLys 357
 Db 195365 ACAGAAAGAGAACTCTGTTAAAGAACTGCTCTGTAATTTTCAAAAGTCAAAACAAAA 195306
 QY 358 ||||
 Db 195305 ACT 195303

RESULT 23

LOCUS COS52685 60 bp DNA linear PAT 30-JAN-2004
 DEFINITION Sequence 22320 from Patent W00210449.
 ACCESSION COS52685
 VERSION COS52685.1 GI:41519112
 KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 Shoshan,A., Wasserman,A., Mintz,E., Mintz,L. and Faigler,S.
 Oligonucleotide library for detecting rna transcripts and splice
 variants that populate a transcriptome
 Patent: WO 0210449-A 22320 07-FEB-2002;
 Compen Inc. (US)

FEATURES

Source

Location/Qualifiers
 1. 60
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
 Pred. No.: 2.97e-09 Length: 60
 Score: 19.00 Matches: 19
 Percent Similarity: 100.00% Conserved: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.09% Indels: 0
 Gaps: 0

US-10-054-935-2 (1-614) x COS52685 (1-60)

QY

376 ValCysLysArgGluLeuAArgSerGlnGluThrProGluLysProAArgSerVal 394

Db

3 GTTGTAAAGTGAATTGAGAGAGCAAGAAACCCAGAAAGCCCGCTCTTCAGTG 59

RESULT 24

AC148580 169993 bp DNA linear HTG 05-OCT-2004
 Gasterosteus aculeatus clone ch213-205p22, WORKING DRAFT SEQUENCE.
 7 ordered pieces.
 AC148580
 AC148580.10 GI:53793740
 HTG: HTGS PHASE2; HTGS DRAFT.
 Gasterosteus aculeatus (three spined stickleback)
 Gasterosteus aculeatus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
 Gasterosteidae; Gasterosteus.
 1 (bases 1 to 169993)
 Qu,C., Lin,S., Postlethwait,J. and Roe,B.A.
 Gasterosteus aculeatus BAC Clone ch213-205p22
 Unpublished
 2 (bases 1 to 169993)
 Qu,C., Lin,S., Postlethwait,J. and Roe,B.A.
 Direct Submission
 Submitted (20-MAR-2004) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 3 (bases 1 to 169993)
 Qu,C., Lin,S., Postlethwait,J. and Roe,B.A.
 Direct Submission
 Submitted (05-OCT-2004) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 TITLE
 AUTHORS
 JOURNAL

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 TITLE
 AUTHORS
 JOURNAL
 COMMENT
 On Oct 5, 2004 this sequence version replaced gi:52782497.
 ----- Genome Center
 Center: Department Of Chemistry And Biochemistry
 The University Of Oklahoma
 Center code:UOKNOR

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced

* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 31093: contig of 31093 bp in length
* 31094 31193: gap of unknown length
* 31194 44796: contig of 13603 bp in length
* 44797 44896: gap of unknown length
* 44897 95056: contig of 50160 bp in length
* 95057 95156: gap of unknown length
* 95157 127376: contig of 32220 bp in length
* 127377 127476: gap of unknown length
* 127477 158428: contig of 30952 bp in length
* 158429 158528: gap of unknown length
* 158529 161675: contig of 3147 bp in length
* 161676 161775: gap of unknown length
* 161776 169993: contig of 8218 bp in length.
Location/Qualifiers
1. .169993
/organism="Gasterosteus aculeatus"
/mol_type="genomic DNA"
/db_xref="taxon:69293"
/clone_lib="Gasterosteus aculeatus BAC library chori-213"

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.: 3,81e-05 Length: 169993
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.93% Indels: 0
DB: 2 Gaps: 0

US-10-054-935-2 (1-614) x AC148580 (1-169993)

Qy 224 GlnlleuleuleuGlnleuAspLeuIleGluGlnGlnGlnGlnleuGln 241

Db 76898 CAGATCTCTCTCTCCACATGACCTGATTCAGAGACAGCAGCAGCTGCAG 76951

RESULT 25
BD018136 903 bp DNA linear PAT 27-AUG-2002
LOCUS Novel gene and novel gene fragment cloned in human neuroblastoma.
DEFINITION BD018136
ACCESSION BD018136
VERSION BD018136.1 GI:22559312
KEYWORDS JP 2001245671-A/374.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 903)

REFERENCE
AUTHORS Nakagawara, A.
TITLE Novel gene and novel gene fragment cloned in human neuroblastoma
JOURNAL Patent: JP 2001245671-A 374 11-SEP-2001;
CHIBA PREF. HISAMITSU PHARMACEUTICAL CO INC

COMMENT OS Homo sapiens (human)

PN JP 2001245671-A/374
PD 11-SEP-2001
PR 07-MAR-2000 JP 2000159195
PI AKIRA NAKAGAWARA
PC C12N15/09, C12Q1/68, G01N33/53, G01N33/566// (C12Q1/68, C12R1.91),
PC C12N15/00
CC Novel gene and novel gene fragment cloned in human CC
neuroblastoma
FH Key Location/Qualifiers
FT source 1. .903
/organism="Homo sapiens (human)"

FEATURES

source

Location/Qualifiers
1. .903
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 3,07e-06 Length: 903
Score: 17.00 Matches: 30
Percent Similarity: 96.77% Conservative: 0
Best Local Similarity: 96.77% Mismatches: 0
Query Match: 2.77% Indels: 1
DB: 6 Gaps: 0

US-10-054-935-2 (1-614) x BD018136 (1-903)

Qy 585 GlnAsnPhgGluLeuProTyrLeuAspGluArgSerArgCysArgLeu-GluIleGlnly 604

Db 69 CAAATTTTGAGTACCTCGTTGGATGAGCGGATGAGATTTGGAGATCCAGAA 128

Qy 604 slysgIntnrProHisArgThrCysArgLys 614
Db 129 GAAGCAACACCTCACCAGCGGTGTAGGAAA 159
RESULT 26
BD098074 903 bp DNA linear PAT 27-AUG-2002
LOCUS Novel genes cloned in humanneuroblastoma and fragments thereof.
DEFINITION BD098074
ACCESSION BD098074.1 GI:22643648
VERSION WO 0166719-A/374.
KEYWORDS WO 0166719-A/374.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 903)

REFERENCE
AUTHORS Nakagawara, A.
TITLE Novel genes cloned in humanneuroblastoma and fragments thereof
JOURNAL Patent: WO 0166719-A 374 13-SEP-2001;
CHIBA PREF. HISAMITSU PHARMACEUTICAL CO INC, AKIRA NAKAGAWARA

COMMENT OS Homo sapiens (human)
PN WO 0166719-A/374
PD 13-SEP-2001
PR 02-MAR-2001 WO 2001JP001629
PR 07-MAR-2000 JP 00P 159195
PI AKIRA NAKAGAWARA
PC C12N15/11, C12Q1/68, G01N33/53, G01N33/566
CC Novel genes cloned in humanneuroblastoma and fragments thereof
FH Key Location/Qualifiers
FT source 1. .903
/organism="Homo sapiens (human)"

FEATURES

source

Location/Qualifiers
1. .903
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 3,07e-06 Length: 903
Score: 17.00 Matches: 30
Percent Similarity: 96.77% Conservative: 0
Best Local Similarity: 96.77% Mismatches: 0
Query Match: 2.77% Indels: 1
DB: 6 Gaps: 0

US-10-054-935-2 (1-614) x BD098074 (1-903)

Qy 585 GlnAsnPhgGluLeuProTyrLeuAspGluArgSerArgCysArgLeu-GluIleGlnly 604

Db 69 CAAATTTTGAGTACCTCGTTGGATGAGCGGATGAGATTTGGAGATCCAGAA 128

Qy 604 slysgIntnrProHisArgThrCysArgLys 614
Db 129 GAAGCAACACCTCACCAGCGGTGTAGGAAA 159
RESULT 27
AL929039/c 162385 bp DNA linear VRT 12-NOV-2002

LOCUS AL929039 162385 bp DNA linear VRT 12-NOV-2002

This Page Blank (uspto)

Db	378	CCCTTGCCCTTC	CCGCCCCCGCCACCAAGCAACCGGCACTTGGGGGGGAGACCTGCGCA	437
Oy	121	AlaGlyAlaGlyCySerProArgProIleTyrGlnAlaValLeuProIleGlnThrGly		140
Db	438	GCCGAGCGCGCTGAGCGCCCCCGGCCCAAGTACAGGCGGTGTGTGCCATTCAAGCGGAC		497
Oy	141	SerLeuValAlaAlaAlaIleGlyLeuProThrProTrpAlaGlyAspIleGlyValAlaAla		160
Db	498	TCTCTCGGGGGGGCCCAAGACCTTACGCTCGGGCTGGAGCACAAGGGTGGGGCGGCC		557
Oy	161	SerProAlaAlaThrAlaSerAspProAlaGlyProProPoleuProLeuProGlyPro		180
Db	558	TCCCCCGGTGCCACCGCTCCGACCAGCGGGAGACCCCAACACTACTCTCCCGGGCGC		617
Oy	181	ProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlnGlyValTyrTrpLys		200
Db	618	CCACCCCTCGCGCCACCGCCACCGCGCGGACCTTGCGGCGCCAGAGAGGAGATGGAG		677
Oy	201	SerMetArgLysSerProLeuGlyValGlyValGlySerGlyAlaSerSerGlnAlaAla		220
Db	678	AGTATGAGGAAGAGCGCTCTCGGGGTGTGTGGCGGCTCGGAGAGCTCCAGTCAAGCGCGC		737
Oy	221	CysLeuLysGlnIleLeuLeuLeuGlnLeuAspLeuIleGlnGlnGlnGlnGlnLeu		240
Db	738	TGCTCTCAACAGATCTCTTCTGTGCAATTGAACTCTCATCGAACACAGACAGACAGCTG		797
Oy	241	GlnAlaLysGlyLysGlyLeuIleGlnLeuLeuLysSerGluArgAspThrLeuLeuAlaArg		260
Db	798	CAGGCGCAAGAAAGAGATCGAGGAGCTGAAGTCAGAGAGAGACAGCTCTTGCTCGG		857
Oy	261	IleGluArgMetGluArgArgMetGlnLeuValLysLysAspAsnGluLysGluArgHis		280
Db	858	ATTGAACCTATGGAAGGAGCGATGCACTGTGTAAAGAAAGGATACGAAAGAAAGGAC		917
Oy	281	LysLeuPheGlnGlyTyrGluThrGlnGlnLysArgGlnGluThrGlnLeuSerGlnLysIle		300
Db	918	AAGCTGTTCAGGGCTATGAACTGAAAGAGAGAGGAAACAGACTATCTGAAGAAATT		977
Oy	301	LysLeuGlnCyGlnProGlnLeuSerGlnThrSerGlnThrLeuProProLysProPhe		320
Db	978	AAACTGGAGTCCACGCGGAGCTTCCGAGACATCCAGACTCTGCTCCCAAGCCCTTC		1037
Oy	321	SerCysGlyArgSerGlyLysGlyValIleLysArgLysSerProPheGlySerThrGluArg		340
Db	1038	TCATGTGGCGCGAGTGAAGAGGACATTAAGAAATCCCATTTGGAAGTACAGAAAGA		1097
Oy	341	LysThrProValLysLysLeuAlaProGlnPheSerLysValLysThrLysThrProLys		360
Db	1098	AAGACTCTCTGTTAAAAAGCTGTCTCTTAATTTCTAAAGTCAAAACAAAATCTCTAAG		1157
Oy	361	HisSerProIleLysGlnGluProCysGlySerLeuSerGluThrValCysLysArgGlu		380
Db	1158	CACCTCTCTATTAAAGAGAAACCTGTGTCTCTTATCTGAACCTGTTGTAAACGTGA		1217
Oy	381	LeuArgSerGlnGluThrProGlnLysProArgSerSerValAspThrProProArgLeu		400
Db	1218	TTGAGAGACCAAGAAACCCCAAGAAAGCCCGGTCTTCAGTGGACACCCCAACAAAGCTC		1277
Oy	401	SerThrProGlnLysGlyLysProSerThrHisProLysGlnLysAlaPheSerSerGluIle		420
Db	1278	TTCACCTCCCAAAAGAGACCCACGACCCATCCCAAGAGAAAGCCCTTCTCAAGTGAAGTA		1337
Oy	421	GluAspLeuProLysLeuSerThrThrGlnMetTyrLeuCysArgTyrPheGlnProPro		440
Db	1338	GAAAGTTTGGCGTACCTTTCCACCAAGAAATGTATTGTGTGTGTGGACACAGCTTCC		1397
Oy	441	ProSerProLeuProLeuArgLysSerSerProLysLysGlnGluThrValAlaArgCys		460
Db	1398	CCATCACCGTACCACTTACGGGAATCTCTCCAAAGAAAGAGAGACGTGTGACAAAGTGT		1457
Oy	461	LeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrpArgAsp		480
Db	1458	CTGAAGCATCAAGTGTGACAGGAAGAACTTCAAGTCTTGCGTGTCTTCTTGGAGGAGAC		1517

Qy	461	HisSerValIGluProLeuArgAspProAsnProSerAspLeuLeuGluAsnLeuAspAsp	500
Db	1518	CACTCAGTAGAGCCCTCTAAGGAGCCCAATCCTTCAGACCTTTTGGAGAACCTCGATGAC	157
Qy	501	SerValPheSerLeuArgHisGAlaIysLeuGluLeuAspGluIysArgArgIysArgTrp	520
Db	1578	ACTGCTTTTCGAGACCGGCATGCACAACTGGAGCTCGATGAGAGAGAGAGAAAGATGG	163
Qy	521	AspIleGlnArgIleArgGluGlnArgIleLeuGlnArgLeuGlnLeuArgMetYrIys	540
Db	1638	GATATTTCAGAGGATCAGGGAAACAAAGAAATTTTACAGCGACTCGAGCTCAGAAATGTATAA	169
Qy	541	LysIysGlyIleGlnGluSerGluProGluValIThrSerPheProGluProAspAsp	560
Db	1698	AAGAAAGAAATTCAGAAATTCGAGCTGAGGTTCACCTATTTTCCCTAGCCAGATGAT	175
Qy	561	ValGluSerLeuMetIleThrProPheLeuProValIAlaIlePheGlyArgProLeuPro	580
Db	1758	GTTGAAAGTTGATGATTAACCCCTCTTCTGCTGTGTAGCATTTGGAGAGCATTACCA	181
Qy	581	LysLeuThrProGlnAsnPheGluLeuProTrpLeuAspGluArgSerArgCysArgLeu	600
Db	1818	AAATTAACCTCCACAGAAATTTTGACTACCTCGTTGGATGAGACGTGACCGATGCAGATTG	187
Qy	601	GluIleGlnIysGlnIThrProHisArgThrCysArgIys	614
Db	1878	GAGATCCAGAGAGAGAACACCTCACCGAGCTGAGAAA	1919
RESULT 2			
US-10-144-194A-65			
/ Sequence 65, Application US/10144194A			
/ Publication No. US20030215809A1			
/ GENERAL INFORMATION:			
/ APPLICANT: Origene Technologies Inc			
/ TITLE OF INVENTION: Regulated Breast Cancer Genes			
/ FILE REFERENCE: 3U 103 R1			
/ CURRENT APPLICATION NUMBER: US/10/144, 194A			
/ CURRENT FILING DATE: 2002-06-12			
/ NUMBER OF SEQ ID NOS: 114			
/ SOFTWARE: PatentIn version 3.0			
/ SEQ ID NO 65			
/ LENGTH: 4372			
/ TYPE: DNA			
/ ORGANISM: Homo sapiens			
/ FEATURE:			
/ NAME/KEY: CDS			
/ LOCATION: (78) ..(1922)			
US-10-144-194A-65			
Alignment Scores:			
Pred. No.: 0			
Score: 614.00			
Percent Similarity: 100.00%			
Best Local Similarity: 100.00%			
Query Match: 100.00%			
DB: 17			
Gaps: 0			
US-10-054-935-2 (1-614) x US-10-144-194A-65 (1-4372)			
Qy	1	MetThrMetArgSerAlaValPheIysAlaAlaAlaProAlaGlyIysProGlu	20
Db	78	ATGACCATAGATCCGGGTGTTCAAAGCGCGCGGCCCTCGCGCGGCAATCTGAG	117
Qy	21	GlnArgLeuAspTrpGluArgAlaAlaAlaLeuGlyIleProGluAspGluProGlyAla	40
Db	138	CAGCACTGACTACGAGCGGGCTGGCGGCTGGGGCGGCCGAGCAGAGCTGGGGCG	197
Qy	41	AlaGlnAlaHisPheLeuProArgHisArgIysIleIysGlnIProGlyProProLeuAla	60
Db	198	GCGGAAGCCACTCTCTCCCGGCAACGTAAGCTCAAGAGACCCGGGCCCTCGGCC	257
Qy	61	SerSerGlnGlyIysSerProAlaProSerProAlaGlyCysGlyIleIysGlyArgIy	80

QY 21 GlnArgLeuAspTyrGluArgAlaAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
Db 138 CAGCGACTGACATACGAGCGGGCTGCGCGCTGGCGGGCCGACGAGCGAGCTGGGGCG 197
QY 41 AlaGluAlaHisPheLeuProArgHisArgLeuLeuGluProGlyProProLeuAla 60
Db 198 GCGGAAGCCCATCTTCTCCCGCGGACCTGAGCTCAAGAGCGGGGCCCTCCCTGGCC 257
QY 61 SerSerGlnGlySerProAlaProSerProAlaGlyCysGlyGlyGlyArgGly 80
Db 258 TCTCTCCAGGGCGGAGCGCCCGCGCTTCCCGGCGGCTGGCGGACGAGGGCGGGGCG 317
QY 81 LeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGln 100
Db 318 TTCTTCTCCCGCGGCGGGCGGGCGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 377
QY 101 ProLeuProCysProProProAlaThrLeuGlnAlaGlyIleGlyGlyGlyProAlaAla 120
Db 378 CCTTGGCTGTGCA 437
QY 121 AlaGlyAlaGlyCysSerProArgProGlyGlnAlaValLeuProIleGlnThrGly 140
Db 438 GCGGAGCGGGCTGAGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 497
QY 141 SerLeuValAlaAlaAlaAlaGlyGluProThrProTrrAlaGlyAspGlyGlyAlaAla 160
Db 498 TCTCTGCTGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 557
QY 161 SerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeuProGlyPro 180
Db 558 TCCCGCGCTCCACCGCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 617
QY 181 ProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlnGlyArgTrrLys 200
Db 618 CCAACCCCTCG 677
QY 201 SerMetArgLysSerProLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 220
Db 678 AGATATGAGGAAGCGCTCTCGGGGGGTGGCGGCGCTCGGAGCGCTCAAGTCAGGCGCC 737
QY 221 CysLeuLysGlnIleLeuLeuLeuGlnLeuAspLeuIleGlnGlnGlnGlnGlnGln 240
Db 738 TGCCTCAACAGATCTTCTGCTGCAATTGAGCTCTCAACAGCGAGCGAGCGAGCTG 797
QY 241 GlnAlaLysGlnLysGlnIleGlnGlnLysSerGlnArgAspThrLeuLeuAlaArg 260
Db 798 CAGGCCAAGGAAAAAGAGATCGAGAGCTGAGTCAAGAGAGAGAGAGAGAGAGAGAG 857
QY 261 IleGluArgMetGluArgArgMetGlnLeuValLysLysAspAsnGlnLysGlnArgHis 280
Db 858 ATTGAACGATGGAAGCGGATGCACTGTAAGAAAGATTAACGAGAAAGAAAGCGCAC 917
QY 281 LysLeuPheGlnGlnGlyTrrGluThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
Db 918 AACCTGTTTACGAGCTATGAACTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 977
QY 301 LysLeuGlnCysGlnProGlnLeuSerGlnThrSerGlnThrLeuProProLysProPhe 320
Db 978 AAACGTGAGTCCAGCGGAGCTTCCGAGAGCATTCCAGACTCGCTCCCAAGCCTTTC 1037
QY 321 SerCysGlyArgSerGlyGlyGlyHisLysArgLysSerProPheGlySerThrGlnArg 340
Db 1038 TCATGTGGCGGAGTGAAGAGGAGCATAAAGAGAAATCCCATTTGAGAGTACAGAAAG 1097
QY 341 LysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLys 360
Db 1098 AAGATCTCTGTAAAGAGCTGGCTCTGAAATTTTCAAAAGTCAAAAGCAAAATCTCTAAG 1157
QY 361 HisSerProIleLysGlnGluProCysGlySerLeuSerGlnThrValCysLysArgGlu 380
Db 1158 CACTCTCTATTAAAG 1217

QY 381 LeuArgSerGlnGlnThrProGlnLysProArgSerSerValAspThrProProArgLeu 400
Db 1218 TTGAAGAGCCAAAGAACCCCAAGAAAGCCCGGATCTTCACTGAGACACCCCAAGACTC 1277
QY 401 SerThrProGlnLysGlyProSerThrHisProLysGlnLysAlaPheSerSerGlnIle 420
Db 1278 TCCACTTCCCAAAAGGAGCCCAAGACCCATCCCAAGAGAGAAACCTTCTCAATGAGATA 1337
QY 421 GluAspLeuProTrrLysSerThrThrGluMetTrrLeuCysArgTrrHisGlnProPro 440
Db 1338 GAAGATTTGCGTACCTTTCACACAGAAATATTTGTGTGTGGCACCGCTCCCGCC 1397
QY 441 ProSerProLeuProLeuArgGlnSerSerProLysLysGlnGlnThrValAlaArgCys 460
Db 1398 CCATCACCGTTACATTACCGAGATCTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAG 1457
QY 461 LeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrrArgAsp 480
Db 1458 CTGATGCCATCAAGTGTTCAGAGAGAAACTTCAGTCTTGGCTTCTTCTTGGAGGGAC 1517
QY 481 HisSerValGlnProLeuArgAspProAsnProSerAspLeuLeuGlnLysAsnLeuAspAsp 500
Db 1518 CACTCAGTAGAGCCTCAAGGAGACCCCAATCTTCAGACCTTTTGGAGAACTGATGAC 1577
QY 501 SerValPheSerLysArgHisAlaLysLeuGlnLeuAspGlnLysArgLysArgTrr 520
Db 1578 AGTGTGTTTTCGAAGCGGCGATGCAAACTGGAAGTGAAGAGAGAGAGAGAGAGAG 1637
QY 521 AspIleGlnArgIleArgGlnGlnArgIleLeuGlnArgLeuGlnLeuArgMetTrrLys 540
Db 1638 GATTTTCAGAGATTCAGGAGACAAAGATTTTACAGCACTGAGCTCAGAAATGTATAA 1697
QY 541 LysLysGlyIleGlnGlnSerGlnProGlnValThrSerPhePheProGlnProAspAsp 560
Db 1698 AAGAAAGATTCAGAAATTCAGCTGAGCTTACCTCATTTTCCCTGAGCGAGATGAT 1757
QY 561 ValGluSerLeuMetIleThrProPheLeuProValValAlaPheGlyArgProLeuPro 580
Db 1758 GTTGAAGATTGATGATTAACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1817
QY 581 LysLeuThrProGlnAsnPheGlnLeuProTrrLeuAspGlnLysArgSerArgCysArgLeu 600
Db 1818 AAATTAACCTCAAGATTTTGAAGTACCTCGTTGGATGAAGCTGACGATGACGATG 1877
QY 601 GlnIleGlnLysLysGlnThrProHisArgThrCysArgLys 614
Db 1878 GAGATCCAGAAAGAACCACTCACCGGAGCTGTAGGAAA [19]9

RESULT 4
US-09-925-302-212
; Sequence 212, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OR INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 212
; LENGTH: 1529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-302-212

Alignment Scores:
Pred. No.: 2,48e-259 Length: 1529
Score: 278.00 Matches: 278

QY 441 ProSerProLeuProLeuArgGluSerSerProLylySgluGluThrValAla 458
Db 783 CCATCACCGTTACATTACGGAAATCCTCCCAAGAGAGAGACTGATGACA 836

RESULT 6
US-10-641-643-322
Sequence 322, Application US/10641643
Publication No. US20040077003A1
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
Susan G. Stuart
Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
GENE EXPRESSION
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/641,643
FILING DATE: 14-Aug-2003
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 322:
SEQUENCE CHARACTERISTICS:
LENGTH: 1290 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TIVANOR01
CLONE: 140704
SEQUENCE DESCRIPTION: SEQ ID NO: 322 :
US-10-641-643-322

Alignment Scores:
Pred. No.: 4.36e-172 Length: 1290
Score: 188.00 Matches: 201
Percent Similarity: 99.50% Conservative: 0
Best Local Similarity: 99.50% Mismatches: 0
Query Match: 30.62% Indels: 1
DB: 17 Gaps: 0

US-10-054-935-2 (1-614) x US-10-641-643-322 (1-1290)

QY 258 LeuAlaArgTLeuGluArgMetGluArgArgMetGluLeuVallylsysAspAsnGluLys 277
Db 3 CTGGCTCGGATTGAACGATATGGAAGCGGATGCACTGTGAAGAAAGATTAACGAGAAA 62

QY 278 GUAARGHISLYSLEUPHEGINGLYTYRGluThrGluGluArgGluGluThrGluLeuSer 297
Db 63 GAAAGCACAAGACTGTTTCAGGGCTATGAAGAACTGAAGAGAGAGAGAAACAGAGCTATCT 122

QY 298 GlnLysLleLysLeuGluCysGlnProGluLeuSerGluThrSerGlnThrLeuProPro 317

Db 123 GAGAAAATTAACTGAGAGTCCAGCCGACCTTCCGAGACATCCAGACTGCGCTCCC 182
QY 318 LysProPheSerCysGlyArgSer-GlyLysGlyHISLYSARGLYSERProPheGlySe 337
Db 183 AAGCCCTTCTCATGTGGCGGAGTGGAGAAAGGACATAAAGGAAATCCCATTTGGAG 242

QY 337 rThrGluArgLysThrProVallylsLysLeuAlaProGluPheSerLysVallysthrly 357
Db 243 TACGAAAGAAAGACCTCTGTAAAAAGCTGGCTCTGAAATTTTCAAAAGTCAAAACAAA 302

QY 357 sThrProLysHISerProIleLysGluGluProCysGlySerLysSerGluThrValCy 377
Db 303 AACTCTTAAGCATCTTCCTATTAAAGAGAACCTGTGTTCTTATCTAAACTGTTTG 362

QY 377 sLysArgGluLeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrPr 397
Db 363 TAAACGTGAATTGAGAGAGCCAGAAACCCAGAAAGCCCGGCTTCACTGAGACACCCC 422

QY 397 oProArgLysSerThrProGluLysGlyProSerThrHisProLysGluLysAlaPheSe 417
Db 423 ACCAAGACTCTCCACTCCCAAGAGGAGACCCAGACCCATCCCAAGAGAGAAACCTTCTC 482

QY 417 rSerGluIleGluAspLeuProTyrLeuSerThrThrGluMetTyrLeuCysArgTrpH 437
Db 483 AAGTGAGATGGAAGATTGCGCTACCTTCCACACAGAAATGATTGTGTGTGGCA 542

QY 437 sGlnProProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrVa 457
Db 543 CCAGCTCCCGCCATCACCGTTACCATTTACGGGAATCCTCCCAAGAGAGAGACTGT 602

QY 457 lAla 458
Db 603 AGCA 606

RESULT 7
US-09-918-995-30064
Sequence 30064, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30064
LENGTH: 365
TYPE: DNA
ORGANISM: Homo sapiens
US-09-918-995-30064

Alignment Scores:
Pred. No.: 1.19e-104 Length: 365
Score: 118.00 Matches: 118
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.22% Indels: 0
DB: 10 Gaps: 0

US-10-054-935-2 (1-614) x US-09-918-995-30064 (1-365)

QY 282 LeuPheGlnGlyTYRGluThrGluGluArgGluGluThrGluLeuSerGluLysIleLys 301
Db 11 CTGTTTCAGGGCTATGAACTGAAGAGAGAGAGAGAAACAGAGCTATCTGAGAAATTTAAA 70

QY 302 LeuGluCysGlnProGluLeuSerGluThrSerGluThrLeuProProLysProPheSer 321
Db 71 CTGGAGTGCAGCGGAGCTTTCGAGACATCCAGACTGTGCTTCCCAAGCCCTTCTCA 130

```
QY 322 CysGlyArgSerGlyLysIleLysArgLysSerProPheGlySerThrGluArgLys 341
DB 131 TGTCGGCGAGTGGAAAGGACATATAAGAAATCCCATTTGGAAGTACGAAAGAAAG 190
QY 342 ThrProValLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 361
DB 191 ACTCTCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 250
QY 362 SerProLysGlyGluGluProCysGlySerLysSerGluThrValCysLysArgLys 381
DB 251 TCTCTCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 310
QY 382 ArgSerGlnGluThrProGluLysProArgSerSerValAspThrProProArg 399
DB 311 AGGAGCCCAAGAAACCCCAAGAAAGCCCGGCTTCAGTGGACACCCCAAGAGA 364

RESULT 8
US-10-242-535A-31928
; Sequence 31928, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31928
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)..(4)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-31928

Alignment Scores:
Pred. No.: 5,1e-58 Length: 440
Score: 70.00 Matches: 70
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.40% Indels: 0
DB: 17 Gaps: 0

US-10-054-935-2 (1-614) x US-10-242-535A-31928 (1-440)
QY 389 LysProArgSerSerValAspThrProProArgLysSerThrProGlnLysGlyProSer 408
DB 7 AAGCCCCGGTCTTCAGTGACACCCCAAGACTCTCCCAAGAAAGGAGGAGGAGGAGG 66
QY 409 ThrHisProLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 428
DB 67 ACCCATCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 126
QY 429 ThrGluMetLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 448
DB 127 ACAGAAATGATTTGTGTGTCGTTGGCACACGCTCCCATACCGTTACATTACGGGAA 186
QY 449 SerSerProLysLysGluGluThrValAla 458
DB 187 TCCTCTCCAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 216
```

```
RESULT 9
US-10-085-783A-31928
; Sequence 31928, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31928
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)..(4)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-31928

Alignment Scores:
Pred. No.: 5,1e-58 Length: 440
Score: 70.00 Matches: 70
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.40% Indels: 0
DB: 17 Gaps: 0

US-10-054-935-2 (1-614) x US-10-085-783A-31928 (1-440)
QY 389 LysProArgSerSerValAspThrProProArgLysSerThrProGlnLysGlyProSer 408
DB 7 AAGCCCCGGTCTTCAGTGACACCCCAAGACTCTCCCAAGAAAGGAGGAGGAGGAGG 66
QY 409 ThrHisProLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 428
DB 67 ACCCATCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 126
QY 429 ThrGluMetLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 448
DB 127 ACAGAAATGATTTGTGTGTCGTTGGCACACGCTCCCATACCGTTACATTACGGGAA 186
QY 449 SerSerProLysLysGluGluThrValAla 458
DB 187 TCCTCTCCAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 216

RESULT 10
US-09-917-800A-725
; Sequence 725, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
```


CC conditions relating to breast cancer, and may be used in gene therapy or
CC antisense therapy. They can also be used in research, drug discovery,
CC clinical medicine and forensic medicine. Sequences given in records
CC ACC72012-ACC72074 represent polynucleotides of the invention that are
CC differentially-regulated in breast cancer. NOTE: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pcc_sequences

XX Sequence 4372 BP; 1136 A; 1012 C; 1077 G; 1147 T; 0 U; 0 Other;

Query Match 100.0%; Score 4372; DB 8; Length 4372;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGTCTCGACCCCGGACCTGCGCCCTTCCACCCCTCTCCGCTCGGTGCGCG 60
DB 1 CAGTCTCGACCCCGGACCTGCGCCCTTCCACCCCTCTCCGCTCGGTGCGCG 60
QY 61 CGCTGCTCCGACCACTATGACATGAGATCCGGGTGTTCAAGCGCGCGCCCTG 120
DB 61 CGCTGCTCCGACCACTATGACATGAGATCCGGGTGTTCAAGCGCGCGCCCTG 120
QY 121 CGGGGGGCAATCCTGAGCGACCTGAGCTAGAGGGGCTGCGGGCTGCGGGCCCG 180
DB 121 CGGGGGGCAATCCTGAGCGACCTGAGCTAGAGGGGCTGCGGGCTGCGGGCCCG 180
QY 181 AGAGCGAGCCTGGGGCGGCGGAGCCCACTTCTCCCGGCAACGTAAAGCTCAAGAGC 240
DB 181 AGAGCGAGCCTGGGGCGGCGGAGCCCACTTCTCCCGGCAACGTAAAGCTCAAGAGC 240
QY 241 CGGGGGCCCCGCTGCTCTCTCCAGGGGGGAGCCCCCGCTTCCCGGGCGGCTGCG 300
DB 241 CGGGGGCCCCGCTGCTCTCTCCAGGGGGGAGCCCCCGCTTCCCGGGCGGCTGCG 300
QY 301 GGGGCAAGGGCGGGGCTTGTACTCCGGGCGGGGGGCGCCCGGGGAGAGAGAGA 360
DB 301 GGGGCAAGGGCGGGGCTTGTACTCCGGGCGGGGGGCGCCCGGGGAGAGAGAGA 360
QY 361 GGTGGGGCGGTTTCTGATGCTTGTGCTGCTCCGCGCGGGCGGCAAGAGCGGAGTTG 420
DB 361 GGTGGGGCGGTTTCTGATGCTTGTGCTGCTCCGCGCGGGCGGCAAGAGCGGAGTTG 420
QY 421 GGGGGAGGCTTGGCGGCGGAGCGGCGGCTGCAAGCCCCGGGCAAGATCAGGCGGTG 480
DB 421 GGGGGAGGCTTGGCGGCGGAGCGGCGGCTGCAAGCCCCGGGCAAGATCAGGCGGTG 480
QY 481 TGGCCATTGAGCGGGCTCTCTGTTGGCGGGCGGCAAGAGCCTTAGCCCTGGGCTGGG 540
DB 481 TGGCCATTGAGCGGGCTCTCTGTTGGCGGGCGGCAAGAGCCTTAGCCCTGGGCTGGG 540
QY 541 ACAAGGTTGGGGCGGCTTCCCGCGCTGCCACCGCTCGGACCGGGGGGACCCCAAC 600
DB 541 ACAAGGTTGGGGCGGCTTCCCGCGCTGCCACCGCTCGGACCGGGGGGACCCCAAC 600
QY 601 TACCTTGTGCGGGCGGCAACCCCTTGGCGGCAACCGCGGAGCCCTTGGCGGCA 660
DB 601 TACCTTGTGCGGGCGGCAACCCCTTGGCGGCAACCGCGGAGCCCTTGGCGGCA 660
QY 661 GGGAGGGGAGATGGAAGATGAGAGAGAGCCCTTCCGGGGGTGGTGGGCTCGGAG 720
DB 661 GGGAGGGGAGATGGAAGATGAGAGAGAGCCCTTCCGGGGGTGGTGGGCTCGGAG 720
QY 721 CCTCCAGTCAAGGCGGCTGCTCAACAGATCTTCTGCTGCAATTGSACTCATGSAAC 780
DB 721 CCTCCAGTCAAGGCGGCTGCTCAACAGATCTTCTGCTGCAATTGSACTCATGSAAC 780
QY 781 AGCAGCAGCAGCAGCTGCAAGGCCAAGAGAAAGAGATCGAGAGCTGAATCGAGAGAG 840
DB 781 AGCAGCAGCAGCAGCTGCAAGGCCAAGAGAAAGAGATCGAGAGCTGAATCGAGAGAG 840
QY 841 AACGCTCTTGTGCTCGGATTTGAACGTATGAAAGGGGATGCAAGCTGTGTAAGAGATG 900
DB 841 AACGCTCTTGTGCTCGGATTTGAACGTATGAAAGGGGATGCAAGCTGTGTAAGAGATG 900

DB 841 AACGCTCTTGTGCTCGGATTTGAACGTATGAAAGGGGATGCAAGCTGTGTAAGAGATG 900
QY 901 ACAGAGAAAGAGCAAGAGCTGTTTCAGGCTATGAACTGAGAGAGAGAGAAACAG 960
DB 901 ACAGAGAAAGAGCAAGAGCTGTTTCAGGCTATGAACTGAGAGAGAGAGAAACAG 960
QY 961 AGCTATCTGAGAAATTTAACTGAGAGTGCACCGGAGCTTTCGAGACATCCGACATC 1020
DB 961 AGCTATCTGAGAAATTTAACTGAGAGTGCACCGGAGCTTTCGAGACATCCGACATC 1020
QY 1021 TGCCTCCCAAGCCCTTCTCATGTGGGCGGAGTGAAGGAGCATTAAGAAATCCCAT 1080
DB 1021 TGCCTCCCAAGCCCTTCTCATGTGGGCGGAGTGAAGGAGCATTAAGAAATCCCAT 1080
QY 1081 TTGGAAGTACAGAAAGAAAGACTCTGTTTAAAGCTGCTCTGAATTTTCAAAAGTCA 1140
DB 1081 TTGGAAGTACAGAAAGAAAGACTCTGTTTAAAGCTGCTCTGAATTTTCAAAAGTCA 1140
QY 1141 AAACAAAAATCTCTAAGCACTCTCTATTTAAAGAGAAACCTGTGGTCTTTATCTGAAA 1200
DB 1141 AAACAAAAATCTCTAAGCACTCTCTATTTAAAGAGAAACCTGTGGTCTTTATCTGAAA 1200
QY 1201 CTGTTTGTAAAGGTGAATTGAGAGCAAGAAACCCAGAAAGCCCGGCTTTCAGTGG 1260
DB 1201 CTGTTTGTAAAGGTGAATTGAGAGCAAGAAACCCAGAAAGCCCGGCTTTCAGTGG 1260
QY 1261 ACAACCCACCAAGACTCTTCACTCCCAAAAGGAGCCAGCAACCATCCCAAGAGAAAG 1320
DB 1261 ACAACCCACCAAGACTCTTCACTCCCAAAAGGAGCCAGCAACCATCCCAAGAGAAAG 1320
QY 1321 CCTTCTCAAGTGAATGAGAAATTTGCGGTACTTTCACCAAGAAATGATTTGTGTG 1380
DB 1321 CCTTCTCAAGTGAATGAGAAATTTGCGGTACTTTCACCAAGAAATGATTTGTGTG 1380
QY 1381 GTTGGCACCAAGCCTCCCATACAGGTTCACATTAACGGAAATCTCTCCAAAGAGAGAG 1440
DB 1381 GTTGGCACCAAGCCTCCCATACAGGTTCACATTAACGGAAATCTCTCCAAAGAGAGAG 1440
QY 1441 AGACTGTAGCAAGGTGTGATGCCATCAAGTGTTCAGAGAGAACTTCAAGTCTTGCGTG 1500
DB 1441 AGACTGTAGCAAGGTGTGATGCCATCAAGTGTTCAGAGAGAACTTCAAGTCTTGCGTG 1500
QY 1501 TTCTTTTGGAGGAGCACTCAATGAGACCTCTAAGGAGCCCAATCTTCAAGACCTTT 1560
DB 1501 TTCTTTTGGAGGAGCACTCAATGAGACCTCTAAGGAGCCCAATCTTCAAGACCTTT 1560
QY 1561 TGGAGAACTGTGATGACATGATGTTTTCGAGCGGATCAAACTGAGAGTGAAGAGA 1620
DB 1561 TGGAGAACTGTGATGACATGATGTTTTCGAGCGGATCAAACTGAGAGTGAAGAGA 1620
QY 1621 AGAGAGAGAAAGATGGATATTTCAGAGGATCAGGAGCAAGAAATTTTACAGCGACTGC 1680
DB 1621 AGAGAGAGAAAGATGGATATTTCAGAGGATCAGGAGCAAGAAATTTTACAGCGACTGC 1680
QY 1681 AGCTCAGATGTATTAAGAAAGAAAGAAATTCAGAAATCTGAGGCTGAGCTTACCTCATTTT 1740
DB 1681 AGCTCAGATGTATTAAGAAAGAAAGAAATTCAGAAATCTGAGGCTGAGCTTACCTCATTTT 1740
QY 1741 TCCCTGAGGCAATATGTTGAAGTTTGAATTAATCCCTCTTCTGCTGTGTGATGACAT 1800
DB 1741 TCCCTGAGGCAATATGTTGAAGTTTGAATTAATCCCTCTTCTGCTGTGTGATGACAT 1800
QY 1801 TTGAGCAGCATTACAAATTAATCTCAAGAAATTTTGAAGTACCTGTTGATGAGC 1860
DB 1801 TTGAGCAGCATTACAAATTAATCTCAAGAAATTTTGAAGTACCTGTTGATGAGC 1860
QY 1861 GTAGCCGATGCAATTTGAGATTCAGAGAGCAAAACCTTCAACGAGCGTGTAGAGAAAT 1920
DB 1861 GTAGCCGATGCAATTTGAGATTCAGAGAGCAAAACCTTCAACGAGCGTGTAGAGAAAT 1920
QY 1921 AGCTGTGCTGGCAAGAACCTGTCTTCAATAGTTGTAGCATGCAATTCACGAGATGGC 1980
DB 1921 AGCTGTGCTGGCAAGAACCTGTCTTCAATAGTTGTAGCATGCAATTCACGAGATGGC 1980

QY 1981 AGAGACCTGATATGACCTTGTCTCATATGTATCACTCGCTGATAATACCCTT 2040
DB 1981 AGAGACCTGATATGACCTTGTCTCATATGTATCACTCGCTGATAATACCCTT 2040
QY 2041 TCATACCTCTTACCTTTGTTTTCATTACTCTGATTTACAAAACTCTTCAATTGGCT 2100
DB 2041 TCATACCTCTTACCTTTGTTTTCATTACTCTGATTTACAAAACTCTTCAATTGGCT 2100
QY 2101 AATGTAGATTAGAGGGGATTTGGGATTTCTTTCCCTTTTGGGAAATGGGCTCTC 2160
DB 2101 AATGTAGATTAGAGGGGATTTGGGATTTCTTTCCCTTTTGGGAAATGGGCTCTC 2160
QY 2161 AAGCTAAGCTATAGATGACAGATTCAGAAATTCAGGGGCTGTTTCTATACATTTGC 2220
DB 2161 AAGCTAAGCTATAGATGACAGATTCAGAAATTCAGGGGCTGTTTCTATACATTTGC 2220
QY 2221 CTATGTTAAAGGGGTAAAGGGCTCTCTTCAATAGACATGTGAAGATGAGACGCCCT 2280
DB 2221 CTATGTTAAAGGGGTAAAGGGCTCTCTTCAATAGACATGTGAAGATGAGACGCCCT 2280
QY 2281 TCCCTTAGAGCTGTGCTGATGAGCACTCTTCAACCCTGTACACCCTCTATAGTGG 2340
DB 2281 TCCCTTAGAGCTGTGCTGATGAGCACTCTTCAACCCTGTACACCCTCTATAGTGG 2340
QY 2341 GTATAGTGAATTTTAAACCTTAATATAAAACAACAACCTCACACCTAGAGCTTTAGACAG 2400
DB 2341 GTATAGTGAATTTTAAACCTTAATATAAAACAACAACCTCACACCTAGAGCTTTAGACAG 2400
QY 2401 AAGAGGAATGACAAGTGAAGCAGATGAAGCAAGCCATCTTCAAGATGAGAAAGACATCG 2460
DB 2401 AAGAGGAATGACAAGTGAAGCAGATGAAGCAAGCCATCTTCAAGATGAGAAAGACATCG 2460
QY 2461 GAGAGTTGGTAACTGCTGTAAGAAAGATGTTGTTCAATTTGAAACTATCTGTGATAC 2520
DB 2461 GAGAGTTGGTAACTGCTGTAAGAAAGATGTTGTTCAATTTGAAACTATCTGTGATAC 2520
QY 2521 AGTCATGTGGGAGGGGATGTTGGCTGTGATTTATTTTTCAGTTAATGAGTAACAATTC 2580
DB 2521 AGTCATGTGGGAGGGGATGTTGGCTGTGATTTATTTTTCAGTTAATGAGTAACAATTC 2580
QY 2581 TTATAGTCTCAAAAACCAAAATCTTTGGAAAGAAAGATGGGGATGGTTAGTTTCAGAAC 2640
DB 2581 TTATAGTCTCAAAAACCAAAATCTTTGGAAAGAAAGATGGGGATGGTTAGTTTCAGAAC 2640
QY 2641 AGTTACAGCTGTAAACAAAGCACTTATGTTGGGATGGCATGCCAAAACCTGTATAAA 2700
DB 2641 AGTTACAGCTGTAAACAAAGCACTTATGTTGGGATGGCATGCCAAAACCTGTATAAA 2700
QY 2701 TGTCTTGTATCATCATCTTCTCAAGTATTCCTTCAATGGGGCTTCAATCCCTTTAGAGA 2760
DB 2701 TGTCTTGTATCATCATCTTCTCAAGTATTCCTTCAATGGGGCTTCAATCCCTTTAGAGA 2760
QY 2761 ACTCTGTGTGTGATAGAGACTTATAGGAGGGTATGGGGAGAGTGTGAAATAGTGTCT 2820
DB 2761 ACTCTGTGTGTGATAGAGACTTATAGGAGGGTATGGGGAGAGTGTGAAATAGTGTCT 2820
QY 2821 TCCCTTGGCTGGCAAAATGTCTATCATCTTGAACAAACAGATGTACTTAAAGAGCTTCTCC 2880
DB 2821 TCCCTTGGCTGGCAAAATGTCTATCATCTTGAACAAACAGATGTACTTAAAGAGCTTCTCC 2880
QY 2881 ATTCACTTTGTAATAAATTTGTATGTATGATCAATCTGTGCTCTCTCCCTCCGTTTGG 2940
DB 2881 ATTCACTTTGTAATAAATTTGTATGTATGATCAATCTGTGCTCTCTCCCTCCGTTTGG 2940
QY 2941 TTAATAATATCAGATAGCACTCCAGAGCACTTGTGTCTCAGTGTAGATCCCTATTAAC 3000
DB 2941 TTAATAATATCAGATAGCACTCCAGAGCACTTGTGTGTCTCAGTGTAGATCCCTATTAAC 3000
QY 3001 TATCTGAAAGAAATAGAGCAAGACCTGTGTCTCAATATATAGGAATTCCTTTCT 3060
DB 3001 TATCTGAAAGAAATAGAGCAAGACCTGTGTCTCAATATATAGGAATTCCTTTCT 3060

QY 3061 TTAATCTTACAGACTATTTGTGTGAAAAACAATAGGGGCTTAATCTCTAGAAAGTAGGGG 3120
DB 3061 TTAATCTTACAGACTATTTGTGTGAAAAACAATAGGGGCTTAATCTCTAGAAAGTAGGGG 3120
QY 3121 CTTTATCCCTTAAAGGAATATGTCCCGAATTAATAGCACTTTTAGAGAGAACCAAG 3180
DB 3121 CTTTATCCCTTAAAGGAATATGTCCCGAATTAATAGCACTTTTAGAGAGAACCAAG 3180
QY 3181 GTATGTAGGGGTGTGTGCTGTGCCCATCAGTGTAGACAGAAAGAGATGGGATACATTG 3240
DB 3181 GTATGTAGGGGTGTGTGCTGTGCCCATCAGTGTAGACAGAAAGAGATGGGATACATTG 3240
QY 3241 TGGGAAGAGAGAAAGTTCTCAGGGGCTCCCACTGTCTAAAGTTTGTGTGAGATGT 3300
DB 3241 TGGGAAGAGAGAAAGTTCTCAGGGGCTCCCACTGTCTAAAGTTTGTGTGAGATGT 3300
QY 3301 GATCTGTGCTCTGTGATTTGACTTTTAAAGAAATTAATCTGGACACATGTAGATTC 3360
DB 3301 GATCTGTGCTCTGTGATTTGACTTTTAAAGAAATTAATCTGGACACATGTAGATTC 3360
QY 3361 TTGATGATCTTGTGCTCTTATTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3420
DB 3361 TTGATGATCTTGTGCTCTTATTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3420
QY 3421 TGGGTTTCAATTTGTAACTTCCATCTGTCTTGAAGAGTGGGCTCTCTATAAGGAACTGC 3480
DB 3421 TGGGTTTCAATTTGTAACTTCCATCTGTCTTGAAGAGTGGGCTCTCTATAAGGAACTGC 3480
QY 3481 TGTAACTTCAATTTGTGACGAAGATGTAGAGAAATAGCACTTAATTCCTAGGGGCTC 3540
DB 3481 TGTAACTTCAATTTGTGACGAAGATGTAGAGAAATAGCACTTAATTCCTAGGGGCTC 3540
QY 3541 TCATCTCACACCTTAAGAGAGAGATTTCTAGAAAACTGGGCGCAGATTTCTTGTGTTCTC 3600
DB 3541 TCATCTCACACCTTAAGAGAGAGATTTCTAGAAAACTGGGCGCAGATTTCTTGTGTTCTC 3600
QY 3601 CATCATTTTAATGTGCAAGGCTGTTCAGTTTCTTACTCTTACTCTTAATGTGATATTTCTTC 3660
DB 3601 CATCATTTTAATGTGCAAGGCTGTTCAGTTTCTTACTCTTACTCTTAATGTGATATTTCTTC 3660
QY 3661 GTACGTGTCCAAAAGAAAAGAAAGAACCCATCACTGTCTCTTGAATTTGTTGATCC 3720
DB 3661 GTACGTGTGTCAAAAAGAAAAGAAAGAACCCATCACTGTGTCTCTTGAATTTGTTGATCC 3720
QY 3721 CTCAGTTTCTTGTATTTTCAAGATGTGTGGGTTCTTAATTTTGGGATAGATTAGCAA 3780
DB 3721 CTCAGTTTCTTGTATTTTCAAGATGTGTGGGTTCTTAATTTTGGGATAGATTAGCAA 3780
QY 3781 ATTTAACCATTTGTGTGTGTGCTTACCCAGGGGACTCCCAAGTTTCTGACTTGAAGTGA 3840
DB 3781 ATTTAACCATTTGTGTGTGTGCTTACCCAGGGGACTCCCAAGTTTCTGACTTGAAGTGA 3840
QY 3841 CTGAGAAATTCACAGAGGTGTATCTGTGSCAGATTTAAGTATCTAATTTCTTGGT 3900
DB 3841 CTGAGAAATTCACAGAGGTGTATCTGTGSCAGATTTAAGTATCTAATTTCTTGGT 3900
QY 3901 CTCCTCTCTCCCTAGAGACCTTATTTATTTATTTGCCCTCTCTTGAAGTATTTCTCTTGG 3960
DB 3901 CTCCTCTCTCCCTAGAGACCTTATTTATTTATTTGCCCTCTCTTGAAGTATTTCTCTTGG 3960
QY 3961 ATTTGACTTTGTGAGAGAGGTGGACAGTATGAGCAAGTTTCAAGTCCAAATTT 4020
DB 3961 ATTTGACTTTGTGAGAGAGGTGGACAGTATGAGCAAGTTTCAAGTCCAAATTT 4020
QY 4021 ACAAGTGTATAGTGTGGGGGAAAAATTAATTTTCCCTACATGGGATACACA 4080
DB 4021 ACAAGTGTATAGTGTGGGGGAAAAATTAATTTTCCCTACATGGGATACACA 4080
QY 4081 CTGTGAATTCATCTTCAACTGAAGGCCCTGTAGTCTCTTAATAACATAGTGTGTTT 4140
DB 4081 CTGTGAATTCATCTTCAACTGAAGGCCCTGTAGTCTCTTAATAACATAGTGTGTTT 4140
QY 4141 TTCTTTAAACAAAGTTTAAGCTAGTGTATTAATAAATAAATAATGCTGTGTCTACT 4200

[illegible]

Db	536	AGGGTGTCTGATGCGCATCAAGTGTTCAGAGAGAACTTCAGTCTTGGCTGTCTTCTTG	595
Qy	1511	GAGGGACCACTCAGTAGAGCCTCTTAAGGACCCAAATCTTCAGACTTTTGGAAACT	1570
Db	596	GAGGGACCACTCAGTAGAGCCTCTTAAGGACCCAAATCTTCAGACTTTTGGAAACT	655
Qy	1571	GGATGACAGTGTGTTTTCGAAGCCGCATTCGAAACTGGAGCTGGATGAGAAAGAAAGAA	1630
Db	656	GGATGACAGTGTGTTTTCGAAGCCGCATTCGAAACTGGAGCTGGATGAGAAAGAAAGAA	715
Qy	1631	AAAGTGGGATTTTACAGAGATCAGGAAACAAAGAAATTTTACAGGACTGCAGCTCAGAT	1690
Db	716	AAAGTGGGATTTTACAGAGATCAGGAAACAAAGAAATTTTACAGGACTGCAGCTCAGAT	775
Qy	1691	GTATTAAGAAAGAAATTCAGAACTGAGCCTGAGGTTACCTCATTTTTCCTGAGCC	1750
Db	776	GTATTAAGAAAGAAATTCAGAACTGAGCCTGAGGTTACCTCATTTTTCCTGAGCC	835
Qy	1751	AGATGATGTTGAAGTTTGATGATTACCCCTTTTCGCTGTGTGATCTTTGGACGACC	1810
Db	836	AGATGATGTTGAAGTTTGATGATTACCCCTTTTCGCTGTGTGATCTTTGGACGACC	895
Qy	1811	ATTACCAAAATTAATCTCCACAGATTTTGTAGCTACCTGGTGGATAGCGTAGCCGATG	1870
Db	896	ATTACCAAAATTAATCTCCACAGATTTTGTAGCTACCTGGTGGATAGCGTAGCCGATG	955
Qy	1871	CAGATTGGAGATCCAGAGAAACCAACACTCACCGAGCGTAGAAATAGCTGTGCTG	1930
Db	956	CAGATTGGAGATCCAGAGAAACCAACACTCACCGAGCGTAGAAATAGCTGTGCTG	1015
Qy	1931	GCAAGAAACCTGTCTTCAGATAGTTGTGATGCGCATTTCCGAGATGGCAGACACTGT	1990
Db	1016	GCAAGAAACCTGTCTTCAGATAGTTGTGATGCGCATTTCCGAGATGGCAGACACTGT	1075
Qy	1991	ATATGTGACCTTTTGCCACATATGTTATACCTGGCTGATTAATACCTTTCATACTCC	2050
Db	1076	ATATGTGACCTTTTGCCACATATGTTATACCTGGCTGATTAATACCTTTCATACTCC	1135
Qy	2051	TTGACTTGTGTTTCATTACTCTGATTTTCACAAAAACCTTTCATTGCGCTAATTTGTAGT	2110
Db	1136	TTGACTTGTGTTTCATTACTCTGATTTTCACAAAAACCTTTCATTGCGCTAATTTGTAGT	1195
Qy	2111	TATGAGGGGTGATTTGGGATTTCTTTCCCTTTTGGGAAATGGGCTCTCAAGCTAAAGC	2170
Db	1196	TATGAGGGGTGATTTGGGATTTCTTTCCCTTTTGGGAAATGGGCTCTCAAGCTAAAGC	1255
Qy	2171	TATAGGATGGCAGATTGAGAGTTTCAGGGGCTGTTTCTATAATTTGCCATATGTTAAA	2230
Db	1256	TATAGGATGGCAGATTGAGAGTTTCAGGGGCTGTTTCTATAATTTGCCATATGTTAAA	1315
Qy	2231	GGGGTAAAGGGCTCTCTTCACTTAGACATGTGAGAGATGAAGCAAGCCCTTCTTTAGAG	2290
Db	1316	GGGGTAAAGGGCTCTCTTCACTTAGACATGTGAGAGATGAAGCAAGCCCTTCTTTAGAG	1375
Qy	2291	CTGTGCTGCAATGGCACTCTTCTCAACCTGTGTACACCTCTTATATAGTGGTATATGAT	2350
Db	1376	CTGTGCTGCAATGGCACTCTTCTCAACCTGTGTACACCTCTTATATAGTGGTATATGAT	1435
Qy	2351	TTTTTAACCCCTAAATTAATAACAAACCTCACCTGATGCTTTAGACACAGAGGGAATG	2410
Db	1436	TTTTTAACCCCTAAATTAATAACAAACCTCACCTGATGCTTTAGACACAGAGGGAATG	1495
Qy	2411	ACAAGTAGAGGATGAAGCAAGCAACATCTTCCACAGATGAGAAAACATTCGAGAGTTGGT	2470
Db	1496	ACAAGTAGAGGATGAAGCAAGCAACATCTTCCACAGATGAGAAAACATTCGAGAGTTGGT	1555
Qy	2471	AGATTAAGTGTGAAAAGATAGTTGTTCAATTGAAACTATTTCTGTGATACAGTCATGTGG	2530
Db	1556	AGATTAAGTGTGAAAAGATAGTTGTTCAATTGAAACTATTTCTGTGATACAGTCATGTGG	1615
Qy	2531	GAAAGGATGTTTGGCTGTGATTAATTTTTCAGTTAATGAGATACAAATTTCTTATCTGCTC	2590
Db	1616	GAAAGGATGTTTGGCTGTGATTAATTTTTCAGTTAATGAGATACAAATTTCTTATCTGCTC	1675

[illegible]

OY		3671	CAAAAAGAAAAAACCACATCAGTGTCTGTGACTTTGGTCTTGANCCCTCAGATTCT	3730
Dd		2756	CAAAAAGAAAAAACCCATCACGTGTCCTTGACTTTGGTCTTGANCCCTCAGATTCT	2815
OY		3731	TCTTGATTTACAGCATGTGTGGGGTTTCCTAATTTTTGGGTATGATTTAGCAAAATTTAACAT	3790
Dd		2816	TCTTGATTTACAGCATGTGTGGGGTTTCCTAATTTTTGGGTATGATTTAGCAAAATTTAACAT	2875
OY		3791	TGTGTTTGTGCCTTACCOCAGGGGACTGCCAGTTTCTGACTTGGAATGACTGAGAAGAA	3850
Dd		2876	TGTGTTTGTGCCTTACCOCAGGGGACTGCCAGTTTCTGACTTGGAATGACTGAGAAGAA	2935
OY		3851	TCCACGAGGTGCTATCTGGCCAGATTAAAGTAGTTTATATTCCTTGSGTCTCCCTCC	3910
Dd		2936	TCCACGAGGTGCTATCTGGCCAGATTAAAGTAGTTTATATTCCTTGSGTCTCCCTCC	2995
OY		3911	CTGAGGACCTCTTAATTTAATTTGTCGCCCTCTCTAGGTTAAATTCCTTGATTTGACTTT	3970
Dd		2996	CTGAGGACCTCTTAATTTAATTTGTCGCCCTCTCTAGGTTAAATTCCTCTTGATTTGACTTT	3055
OY		3971	GTTGAGAGAGAGAGTTGACAGTAGATTAGCAAAGTTCCAAAGTCMAAATTAACAAGTGTGT	4030
Dd		3056	GTTGAGAGAGAGAGTTGACAGTAGATTAGCAAAGTTCCAAAGTCMAAATTAACAAGTGTGT	3115
OY		4031	AGAGTGTGGGGGGAAAAATTAGTCTTATTTTTCCCTACATGGAGATACAACACTGGAATTC	4090
Dd		3116	AGAGTGTGGGGGGAAAAATTAGTCTTATTTTTCCCTACATGGAGATACAACACTGGAATTC	3175
OY		4091	AATCTTCAACTGAAGGCGCTGCAGTCTCCCTAATAACAATAGTGTGTTTTCTTTAACA	4150
Dd		3176	AATCTTCAACTGAAGGCGCTGCAGTCTCCCTAATAACAATAGTGTGTTTTCTTTAACA	3235
OY		4151	AAGTTTAAGCTAGTGTTAATTAATTAATAAAAAAATTCGTGTCTGTCTACTTCAAGCTTGT	4210
Dd		3236	AAGTTTAAGCTAGTGTTAATTAATTAATAAAAAAATTCGTGTCTGTCTACTTCAAGCTTGT	3295
OY		4211	TTTAAGCCCATTTCAATATGTTGTCTGTGTGTGAATTCATPACTTTTGATPCCAATTTCTG	4270
Dd		3296	TTTAAGCCCATTTCAATATGTTGTCTGTGTGTGAATTCATPACTTTTGATPCCAATTTCTG	3355
OY		4271	ATGTGTAAATTTGGTGTCTTGTAAATATCTTAATAAGATTCATTTGTAATTAACATAT	4330
Dd		3356	ATGTGTAAATTTGGTGTCTTGTAAATATCTTAATAAGATTCATTTGTAATTAACATAT	3415
OY		4331	TGTGCTGTGT 4340 	
Dd		3416	TGTGCTGTGT 3425 	
RESULT 3				
AD130841				
ID	AD130841 standard; cDNA, 1545 BP.			
XX				
AC	AD130841;			
DT	17-JUN-2004 (first entry)			
XX				
DE	Human cDNA #167.			
XX				
KM	Human; gene; ss; immunological response; immunopathological condition;			
KM	Crohn's disease; asthma; ulcerative colitis; hyper eosinophilia;			
KM	irritable bowel syndrome; osteoarthritis; rheumatoid arthritis;			
KM	acute monocytic leukaemia; anti-inflammatory; antislathmatic; antiulcer;			
KM	osteopathic; antidiarrhtic; antipneumatic; cyrostatic.			
XX				
OS	Homo sapiens.			
PN	US6607879-B1.			
XX				
PD	19-AUG-2003.			
XX				
PF	09-FEB-1998; 98US-00023655.			
XX				

Db 1441 TGGTGTGTAATCATTAACCTTTGATACCAATTCGTGATGTAATAATGGTGTCTGTGA 1500
 Qy 4296 ATATCTTATAAGAGTTCAATGTGTAATAAATCAATGTGGCTGTT 4340
 Db 1501 ATATCTTATAAGAGTTCAATGTGTAATAAATCAATGTGGCTGTT 1545

RESULT 4
 AA224879
 ID AA224879 standard; DNA; 1509 BP.
 XX AA224879;
 AC
 XX
 DT 02-DEC-1999 (first entry)
 XX
 DE Human secreted protein gene 69 clone HCEK38.
 XX
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asplenia; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; resection; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 PN WO947540-A1;
 XX
 PD 23-SEP-1999.
 XX
 PF 18-MAR-1999; 99WO-US005804.
 XX
 PR 19-MAR-1998; 98US-0078563P.
 PR 19-MAR-1998; 98US-0078566P.
 PR 19-MAR-1998; 98US-0078573P.
 PR 19-MAR-1998; 98US-0078574P.
 PR 19-MAR-1998; 98US-0078576P.
 PR 19-MAR-1998; 98US-0078577P.
 PR 19-MAR-1998; 98US-0078578P.
 PR 19-MAR-1998; 98US-0078579P.
 PR 19-MAR-1998; 98US-0078581P.
 PR 01-APR-1998; 98US-0080312P.
 PR 01-APR-1998; 98US-0080313P.
 PR 01-APR-1998; 98US-0080314P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Ni J, Rosen CA, Yu G, Young PB, Feng P, Soppet DR,
 PI Wei Y, Endress GA, Duan RD, Kyaw H, Edner R, Lafleur DW, Olsen HS;
 PI Shi Y, Moore PA;
 XX
 DR WPI; 1999-562050/47.
 P-PSDB; AA241376.
 XX
 PT New isolated human genes, useful for diagnosis and treatment of e.g.
 PT cancers, neurological disorders, immune diseases, inflammation or blood
 PT disorders.
 XX
 PS Claim 1; Page 340; 484pp; English.

CC new polypeptides in a sample or by determining the presence of mutations
 CC in the new polynucleotides. Specific uses are described for each of the
 CC 95 polynucleotides, based on which tissues they are most highly expressed
 CC in (see AA224811 for described uses)
 XX
 SQ Sequence 1509 BP; 397 A; 272 C; 319 G; 521 T; 0 U; 0 Other;

Query Match 33.8%; Score 1478; DB 2; Length 1509;
 Best Local Similarity 99.9%; Pred. No. 1.8e-292;
 Matches 1500; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 2859 GATGTACCTTAAGAGCTTCTCCATTCATTTGTAATAAATATGATGATGATCCATCTT 2918
 Db 9 GATGTACCTTAAGAGCTTCTCCATTCATTTGTAATAAATATGATGATGATCCATCTT 68
 Qy 2919 GGTCCCTCCCCCGCGGTTTTGTTAAATATGAGATAGACCTCCAGGCACTTGGTC 2978
 Db 69 GGTCTCTCCCGCGGTTTTGTTAAATATGAGATAGACCTCCAGGCACTTGGTC 128
 Qy 2979 TCAGTGAAGATCCCTATTACTATCTGAAGAAATAGAGCCAGACCTCTGTCTCA 3038
 Db 129 TCAGTGAAGATCCCTATTACTATCTGAAGAAATAGAGCCAGACCTCTGTCTCA 188
 Qy 3039 AATATATAGAAATGCTTTCTTTAGCTCTTACCTTACGACTATTGTGAAACAGATAGGGGT 3098
 Db 189 AATATATAGAAATGCTTTCTTTAGCTCTTACGACTATTGTGAAACAGATAGGGGT 248
 Qy 3099 CTATATCTCTAAGAGATAGGGGCTTTATTCCTTAAGAAATATGTCCTCAATATATAG 3158
 Db 249 CTATATCTCTAAGAGATAGGGGCTTTATTCCTTAAGAAATATGTCCTCAATATATAG 307
 Qy 3159 CACTTTAGAGAGAGCCAGAGATATGTA-GGGTGTGTGGCTGGCCATCAGTGAAGAC 3217
 Db 308 CACTTTAGAGAGAGCCAGAGATATGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 367
 Qy 3218 GAAGAGAGATGGATATACATTTGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3277
 Db 368 GAAGAGAGATGGATATACATTTGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
 Qy 3278 GCTAAAGTTTTTGTGAGATGTGATCTGTCTCTCGATTTGACTTTTAAAGAAATTA 3337
 Db 428 GCTAAAGTTTTTGTGAGATGTGATCTGTCTCTCGATTTTAAAGAAATTA 487
 Qy 3338 TTCTGGCAGACATGATGATATCTTGATGATCTGCTGCTTATTTCTCTTTGTGT 3397
 Db 488 TTCTGGCAGACATGATGATATCTTGATGATCTGCTGCTTATTTCTCTTTGTGT 547
 Qy 3398 GT 3457
 Db 548 GT 607
 Qy 3458 GGGCTCTCTAATAAGGAGACCTGTGTAACTTATTCAGACAGAGATGTAAGAGAAATA 3517
 Db 608 GGGCTCTCTAATAAGGAGACCTGTGTAACTTATTCAGACAGAGATGTAAGAGAAATA 667
 Qy 3518 GGACTTAATTCACCTAGGGGCTCTCATCTCACACCTTAAGAGAGATTTCTGAAGAAAC 3577
 Db 668 GGACTTAATTCACCTAGGGGCTCTCATCTCACACCTTAAGAGAGATTTCTGAAGAAAC 727
 Qy 3578 TGGGCGAGATTTCTTGTGTCTTCATCATTTTATGTGGAGGCTGTTCAGTTTCTTAC 3637
 Db 728 TGGGCGAGATTTCTTGTGTCTTCATCATTTTATGTGGAGGCTGTTCAGTTTCTTAC 787
 Qy 3638 TCTTACCTATGATATTTCTTGTGATACGTGTCAAAAAGAAAAAGAACCAATCATGTGT 3697
 Db 788 TCTTACCTATGATATTTCTTGTGATACGTGTCAAAAAGAAAAAGAACCAATCATGTGT 847
 Qy 3698 CTTTACCTATGATATTTCTTGTGATACGTGTCTTGTGATGATGATGATGATGATGATG 3757
 Db 848 CTTTACCTATGATATTTCTTGTGATACGTGTCTTGTGATGATGATGATGATGATGATG 907
 Qy 3758 TATTTTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3817

Db 908 TAATTTGGGTATGAGTTAGCAAAATTTAAACATTGTTGTTGTCCTACCCAGGGGACTC 967
Qy 3818 CCCAGTTTCTGACTTGAAGTACGTAGAGAAATCCACGAGGTCTATCTGGCCAGATTT 3877
Db 968 CCCAGTTTCTGACTTGAAGTACGTAGAGAAATCCACGAGGTCTATCTGGCCAGATTT 1027
Qy 3878 AAGTATGATCTATTTCTGTTCTGCTCCCTCCCTGAGAGCCCTTATTTATTTATGTCCTC 3937
Db 1028 AAGTATGATCTATTTCTGTTCTGCTCCCTCCCTGAGAGCCCTTATTTATTTATGTCCTC 1087
Qy 3938 TCTTTAGGTATATTTCTCCTTTGATTTGAATTGTTGAGAGAGAGGTGGACAGTAGATT 3997
Db 1088 TCTTTAGGTATATTTCTCCTTTGATTTGAATTGTTGAGAGAGAGGTGGACAGTAGATT 1147
Qy 3998 AGCAAAATTCGAGTCCAAATTAATTAACGTGTGTAGAGTGTGGGGGAAATTAATGCTTAT 4057
Db 1148 AGCAAAATTCGAGTCCAAATTAATTAACGTGTGTAGAGTGTGGGGGAAATTAATGCTTAT 1207
Qy 4058 TTTTCCCTCATGGGATACACACGTGGAATTCATCTTCACTGAAAGGCCCTGCAAGTTC 4117
Db 1208 TTTTCCCTCATGGGATACACACGTGGAATTCATCTTCACTGAAAGGCCCTGCAAGTTC 1267
Qy 4118 TCTTAAATAGTATGTTGTTGTTTCTTTTAAACAAAGTTTAAAGTATGTTAATAATTA 4177
Db 1268 TCTTAAATAGTATGTTGTTGTTTCTTTTAAACAAAGTTTAAAGTATGTTAATAATTA 1327
Qy 4178 AAAAAATGCTGTGTCTGCTACTTCACTTGTGTTTATGCCCATTCAATATGTTGCTG 4237
Db 1328 AAAAAATGCTGTGTCTGCTACTTCACTTGTGTTTATGCCCATTCAATATGTTGCTG 1387
Qy 4238 TGTGTAAATCAATCTTGTGATACATTTCTGATGTGTAATTTGTTGCTGTTAAAT 4297
Db 1388 TGTGTAAATCAATCTTGTGATACATTTCTGATGTGTAATTTGTTGCTGTTAAAT 1447
Qy 4298 ATCTTAAAGAGTTCATTTGTAATTAATCTATTTGCTGTGTTAAAAA 4357
Db 1448 ATCTTAAAGAGTTCATTTGTAATTAATCTATTTGCTGTGTTAAAAA 1507
Qy 4358 AA 4359
Db 1508 AA 1509

RESULT 5
ADA39700
ID ADA39700 standard; cDNA; 1509 BP.
XX
AC ADA39700;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human secreted protein encoding cDNA.
XX
KW Human; secreted protein; cancer; hyperproliferative disorder;
KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;
KW anaemia; allergic reaction; asthma; cardiovascular disorder;
KW wound healing; cyclostatic; immunosuppressive; nootropic; neuroprotective;
KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;
KW vulnerary; cardiant; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO2002102993-A2.
XX
PD 27-DEC-2002.
XX
PF 19-MAR-2002; 2002WO-US008123.
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;
PI MPI; 2003-175238/17.
DR
XX
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating cancer or other hyperproliferative disorder,
PT asthma, allergies or AIDS.
XX
XX
PS Claim 9; SEQ ID NO 82; 3205pp; English.
XX
XX
CC The invention relates to novel genes ADA39629-ADA4056 and proteins
CC ADA4056-ADA41501 for human secreted proteins, useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
CC fragments, and agonists or antagonists that bind to the polypeptide are
CC useful for preparing a diagnostic or pharmaceutical composition for
CC diagnosing or treating cancer or other hyperproliferative disorder. The
CC polypeptides and nucleic acid molecules are also useful for detecting,
CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC or other hyperproliferative disorders including neoplasms, autoimmune
CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC thrombocytopenia), allergic reactions including asthma or eczema,
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC fungal or viral infections including HIV/AIDS), or wound healing and
CC disorders of epithelial cell proliferation. The nucleic acids are also
CC useful for chromosome identification, radiation hybrid mapping or long-
CC range restriction mapping, as molecular weight markers, or as
CC hybridization or diagnostic probes. The polypeptides and antibodies are
CC useful for providing immunological probes for differential identification
CC of the tissues immunohistochemistry assays. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1509 BP; 397 A; 272 C; 319 G; 521 T; 0 U; 0 Other;
Query Match 33.8%; Score 1478; DB 8; Length 1509;
Best Local Similarity 99.9%; Pred. No. 1.8e-292;
Matches 1500; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
Qy 2859 GATGTACCTAATGAGCTTCCATTCATCTTGTAAATTAATTTGATGATGATCCATCTT 2918
Db 9 GATGTACCTAATGAGCTTCCATTCATCTTGTAAATTAATTTGATGATGATCCATCTT 68
Qy 2919 GGTCTCTCCCTCCCGTTTGTAAATATGAGATAGACATCCAGGSCCACTTGTGTC 2978
Db 69 GGTCTCTCCCTCCCGTTTGTAAATATGAGATAGACATCCAGGSCCACTTGTGTC 128
Qy 2979 TCAATGTAATGATCCCTATTAATCTATCTGAAGAAATAGACCAAGCTTGTGTC 3038
Db 129 TCAATGTAATGATCCCTATTAATCTATCTGAAGAAATAGACCAAGCTTGTGTC 188
Qy 3039 AATATATAGAAATGCTTCTTTAGTCTTCAGGATATATGTTGTGAAACCAATAGGGGT 3098
Db 189 AATATATAGAAATGCTTCTTTAGTCTTCAGGATATATGTTGTGAAACCAATAGGGGT 248
Qy 3099 CTATATCTCTAGAGGAGGCTTTATCTTTAAAGAAATATGCCCCAGATTATTA 3158
Db 249 CTATATCTCTAGAGGAGGCTTTATCTTTAAAGAAATATGCCCCAGATTATTA 307
Qy 3159 CACTTTTAAAGAGGAGGAGGATATGA -GGGTGTGTGCTGGCCCATCAATGATGAGCAC 3217
Db 308 CACTTTTAAAGAGGAGGAGGATATGAAGGATGTGTGGCTGGCCCATCAATGATGAGCAC 367
Qy 3218 GAAAGAGAAATGGATACATTTGTGGGAGAGAGAAAGTTCTCTAGGGGCTCCCACT 3277

QY 2859 GATGACCTAATGAGCTTCTCCATTCGACTTTGTAATAATTTGTAATGTAACATCTT 2918
DB 9 GATGACCTAATGAGCTTCTCCATTCGACTTTGTAATAATTTGTAATGTAACATCTT 68
QY 2919 GGTCTCTCCCTCCCGTGGTGGTAAATATCAGAGTAGCAGCTCCAGGCGCATTTGGTGC 2978
DB 69 GGTCTCTCCCTCCCGTGGTGGTAAATATCAGAGTAGCAGCTCCAGGCGCATTTGGTGC 128
QY 2979 TCAGTGAAGATCCCTATTAACTATCTGAAGGAAAAATAGAGCCAAAGCTCTGGTCTCA 3038
DB 129 TCAGTGAAGATCCCTATTAACTATCTGAAGGAAAAATAGAGCCAAAGCTCTGGTCTCA 188
QY 3039 AATATATAGGAATGGCTTTCTTTAGCTTCAGAGCATTTGTGTGAAGAAAGTAGGGGT 3098
DB 189 AATATATAGGAATGGCTTTCTTTAGCTTCAGAGCATTTGTGTGAAGAAAGTAGGGGT 248
QY 3099 CTAAATCTCCAGAAAGGAGGAGGCTTTTATCCTTAAAGAAATATGTCCTCCAGATTATTA 3158
DB 249 CTAAATCTCCAGAAAGGAGGAGGCTTTTATCCTTAAAGAAATATGTCCTCCAGATTATTA 307
QY 3159 CACTTTTAAAGAGAAAGCCAAAGGATGTA - GGGTGTGTGGCTGGCCCATCAGTGAAGCAC 3217
DB 308 CACTTTTAAAGAGAGCCAAAGGATGTAAGGGGTGTGGCTGGCCCATCAGTGAAGCAC 367
QY 3218 GAAGAGAGATGGGATTAACATTTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3277
DB 368 GAAGAGAGATGGGATTAACATTTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
QY 3278 GCTAAAGTTTGTGTGAGATGTTGATCTGTGCTTCCGATTTGACCTTTAAAGAAATTA 3337
DB 428 GCTAAAGTTTGTGTGAGATGTTGATCTGTGCTTCCGATTTGACCTTTAAAGAAATTA 487
QY 3338 TTCTGGCAGACAGATGATATTTCTTGATGATCTTGCTGCTTATTTCTCTTTTGTGT 3397
DB 488 TTCTGGCAGACAGATGATATTTCTTGATGATCTTGCTGCTTATTTCTCTTTTGTGT 547
QY 3398 GT 3457
DB 548 GT 607
QY 3458 GGGCTCTCTAATAGGGAACCTGCTGTAACTTCATTGACACAGAGATGTAAGAGAAATA 3517
DB 608 GGGCTCTCTAATAGGGAACCTGCTGTAACTTCATTGACACAGAGATGTAAGAGAAATA 667
QY 3518 GGACTTAATTTCCAGAGGGCTCTCATCTTCACACCTTAAAGAGAGATTTCTAGAAAAAC 3577
DB 668 GGACTTAATTTCCAGAGGGCTCTCATCTTCACACCTTAAAGAGAGATTTCTAGAAAAAC 727
QY 3578 TGGGCGAGATTTCTTTGTTCTTCATCATTTTAATGTGCGAGGCTGTTCAGTTTCTTAC 3637
DB 728 TGGGCGAGATTTCTTTGTTCTTCATCATTTTAATGTGCGAGGCTGTTCAGTTTCTTAC 787
QY 3638 TCTTACCTAATGATATTTCTTTGTTAAGTGTAAAGTGTCAAAAAAGAAAAAGAACCCATCATGTGT 3697
DB 788 TCTTACCTAATGATATTTCTTTGTTAAGTGTGTCAAAAAAGAAAAAGAACCCATCATGTGT 847
QY 3698 CTCTGACCTTTGTTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3757
DB 848 CTCTGACCTTTGTTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 907
QY 3758 TAAATTTGGGTATGAGTATGACAAATTTAACCATTTGTTGTGCTTACCCAGGGGATCTC 3817
DB 908 TAAATTTGGGTATGAGTATGACAAATTTAACCATTTGTTGTGCTTACCCAGGGGATCTC 967
QY 3818 CCAGATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3877
DB 968 CCAGATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1027
QY 3878 AAGTATGATTTCTATTTCTGTTCTGCTCCCTCCCTGAGGACCTTATTTATTTATGTCCTC 3937
DB 1028 AAGTATGATTTCTATTTCTGTTCTGCTCCCTCCCTGAGGACCTTATTTATTTATGTCCTC 1087

QY 3938 TCTTCTAGGTTAATTTCTCCTTTGATTTGACTTTGTTGAAGAGAGGTGGAAGATGAT 3997
DB 1088 TCTTCTAGGTTAATTTCTCCTTTGATTTGACTTTGTTGAAGAGAGGTGGAAGATGAT 1147
QY 3998 AGCAAAATTCGAAAGTGCAGAAATTTACAGTGTGTGAGAGTGGGGGAGAAATTAATGCTTAT 4057
DB 1148 AGCAAAATTCGAAAGTGCAGAAATTTACAGTGTGTGAGAGTGGGGGAGAAATTAATGCTTAT 1207
QY 4058 TTTTCCCTCATGAGGATACACACCTGTGAATTCATCTTCAACTGAAGCCCTGCAGTTTC 4117
DB 1208 TTTTCCCTCATGAGGATACACACCTGTGAATTCATCTTCAACTGAAGCCCTGCAGTTTC 1267
QY 4118 TCCCTAAACATAGTGTGTTGTTTCTTTAACAAGATTAAGCTAGTGTAAATTAATTA 4177
DB 1268 TCCCTAAACATAGTGTGTTGTTTCTTTAACAAGATTAAGCTAGTGTAAATTAATTA 1327
QY 4178 AAAAAATTCCTGCTGTCTACTTCAGCTTGTGTTTATGCCATTTCAATATGTTGCTG 4237
DB 1328 AAAAAATTCCTGCTGTCTACTTCAGCTTGTGTTTATGCCATTTCAATATGTTGCTG 1387
QY 4238 TGTGTAATTCATTAATTTGATACATTTCTGATGTGTAATAATGCTGCTTGAAT 4297
DB 1388 TGTGTAATTCATTAATTTGATACATTTCTGATGTGTAATAATGCTGCTTGAAT 1447
QY 4298 ATCTTAAAGAGTCAATTTGTAATAAATAATTTGCGCTGTTAAAAAATTTAAAAA 4357
DB 1448 ATCTTAAAGAGTCAATTTGTAATAAATAATTTGCGCTGTTAAAAAATTTAAAAA 1507
QY 4358 AA 4359
DB 1508 AA 1509

RESULT 7
ADB91097
ID ADB91097 standard; cDNA; 1509 BP.
XX
AC ADB91097;
XX
XX 04-DEC-2003 (first entry)
DT
XX
DE Human secreted protein cDNA #SEQ ID 43.
XX
XX
XX Secreted protein; gene therapy; antidiabetic; diabetes; human; ss.
XX
OS Homo sapiens.
XX
PN MO2003004622-A2.
XX
XX 16-JAN-2003.
PD
XX
XX 19-MAR-2002; 2002MO-US008124.
PF
XX
PR 21-MAR-2001; 2001US-0277340P.
PR 13-UTL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Ruben SM;
XX
XX MPI, 2003-229407/22.
DR
XX
XX Nucleic acid encoding a human secreted protein is useful in diagnosing or
PT treating diabetes or conditions related to diabetes.
XX
XX
XX Claim 9; SEQ ID NO 43; 1537bp; English.
XX
XX The invention relates to isolated nucleic acid molecules ADB91065-
CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
CC ADB91834. Also disclosed is a recombinant vector comprising a
CC polynucleotide of the invention, and a recombinant host cell comprising
CC the recombinant vector. The polypeptide of the invention is useful in

CC identifying a binding partner by contacting the polypeptide with a
 CC binding partner, and determining whether the binding partner increases or
 CC decreases activity of the polypeptide. The polypeptide, polynucleotide,
 CC antibody or its fragment, agonist or antagonist are useful for preparing
 CC a pharmaceutical composition for diagnosing or treating diabetes or
 CC conditions related to diabetes. The present sequence is that of the human
 CC immunoglobulin Fc portion used to generate fusion proteins, increasing
 CC the stability of the fused protein as compared to the secreted protein
 CC only. Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1509 BP; 397 A; 272 C; 319 G; 521 T; 0 U; 0 Other;

Query Match 33.8%; Score 1478; DB 9; Length 1509;
 Best Local Similarity 99.9%; Pred. No. 1.8e-292;
 Matches 1509; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 2859 GATGATCTAATAGAGCTTCTCCATTCACCTTGTAAATAATTTGTATGTACATCTT 2918
 DB 9 GATGATCTAATAGAGCTTCTCCATTCACCTTGTAAATAATTTGTATGTACATCTT 68
 QY 2919 GGTCTCTCCCTCCCTCTTTTGTAAATAATCAGAGATGACCTCCAGGCACTTGTGTC 2978
 DB 69 GGTCTCTCCCTCCCTCTTTTGTAAATAATCAGAGATGACCTCCAGGCACTTGTGTC 128
 QY 2979 TCAGTGAATCCCTATCTATCTGAAGAATAATAGCCCAAGCCCTGTGTCTCA 3038
 DB 129 TCAGTGAATCCCTATCTATCTGAAGAATAATAGCCCAAGCCCTGTGTCTCA 188
 QY 3039 AATATATAGAAATTCCTTTCTTATGCTTCAGAGACTATTTGTGAATAACAAGTGGGCT 3098
 DB 189 AATATATAGAAATTCCTTTCTTATGCTTCAGAGACTATTTGTGAATAACAAGTGGGCT 248
 QY 3099 CTATATCTCTAGAAAGTGGGCTTTTATCTTAAAGAAATATGTCCTCAATTTAG 3158
 DB 249 CTATATCTCTAGAAAGTGGGCTTTTATCTTAAAGAAATATGTCCTCAATTTAG 307
 QY 3159 CACTTTAGAGGAAGCAAGCATGATGTA-GGGGTGTGGCTGGCCCATAGTGGAGCAC 3217
 DB 308 CACTTTAGAGGAAGCAAGCATGATGTAAGGGGTGTGGCTGGCCCATAGTGGAGCAC 367
 QY 3218 GAAGAGAAATGGATACCATTTGTGGAGAGAAAGAAAGTCTCTCAGGGGCTCCCACT 3277
 DB 368 GAAGAGAAATGGATACCATTTGTGGAGAGAAAGAAAGTCTCTCAGGGGCTCCCACT 427
 QY 3278 GCTAAAGTTTTTGTGATGTTGATCTGTGCTTCTGGAATTGACTTTTAAAGAAATTA 3337
 DB 428 GCTAAAGTTTTTGTGATGTTGATCTGTGCTTCTGGAATTGACTTTTAAAGAAATTA 487
 QY 3338 TTCTGGAGAGCAATGATGTAATCTTGTGATGATCTGTGCTTCTTATTTCTCTTTTGTGT 3397
 DB 488 TTCTGGAGAGCAATGATGTAATCTTGTGATGATCTGTGCTTCTTATTTCTCTTTTGTGT 547
 QY 3398 GT 3457
 DB 548 GT 607
 QY 3458 GGGCTCTCTAATAGGAACCTGCTGTAACCTTCAGCAAGAGATGTAAGAGAAATA 3517
 DB 608 GGGCTCTCTAATAGGAACCTGCTGTAACCTTCAGCAAGAGATGTAAGAGAAATA 667
 QY 3518 GGAATTAATCCACTAGGGGCTCATCTCAACCTTAAAGAGAGATTTCTAAGAAAC 3577
 DB 668 GGAATTAATCCACTAGGGGCTCATCTCAACCTTAAAGAGAGATTTCTAAGAAAC 727
 QY 3578 TGGGCCAGATTTTCTTTGTCTCCATATTTAATGTGCGAGGCTGTTCAGTTTCTTAC 3637
 DB 728 TGGGCCAGATTTTCTTTGTCTCCATATTTAATGTGCGAGGCTGTTCAGTTTCTTAC 787
 QY 3638 TCTTACTATGTATGATTTCTTGTAAAGTGTCCAAAAAGAAAAAGAACCAATCACTGT 3697
 DB 788 TCTTACTATGTATGATTTCTTGTAAAGTGTCCAAAAAGAAAAAGAACCAATCACTGT 847

QY 3698 CTCTTACATTTGTCTTTTATGATCCCTAGATTTCTTGTATTCAGCATGTGTCCGTTCC 3757
 DB 848 CTCTTACATTTGTCTTTTATGATCCCTAGATTTCTTGTATTCAGCATGTGTCCGTTCC 907
 QY 3758 TAATTTGGGATAGATGACCAATTTAACCATTGTGTTTGTGCTTACCCAGGACATC 3817
 DB 908 TAATTTGGGATAGATGACCAATTTAACCATTGTGTTTGTGCTTACCCAGGACATC 967
 QY 3818 CCGAGTTTCTGACTTGAAGATGACGTAAGAAATCCAGAGGTCTATCTGAGCAGATT 3877
 DB 968 CCGAGTTTCTGACTTGAAGATGACGTAAGAAATCCAGAGGTCTATCTGAGCAGATT 1027
 QY 3878 AAGTAGATCTATTTCTTGTGTTCTCCCTCCCTGAGACCTTATTTATTTGTCCTCC 3937
 DB 1028 AAGTAGATCTATTTCTTGTGTTCTCCCTCCCTGAGACCTTATTTATTTGTCCTCC 1087
 QY 3938 TCTTTAGGTTAATTTCTCTTATTTGACTTTGTGAGAGAGATGAGCATGATTT 3997
 DB 1088 TCTTTAGGTTAATTTCTCTTATTTGACTTTGTGAGAGAGATGAGCATGATTT 1147
 QY 3998 AGCAAGTCCAGACCAAAATTAACAGTGTAGAGTGTGGGGGAAATTAAGTCTAT 4057
 DB 1148 AGCAAGTCCAGACCAAAATTAACAGTGTAGAGTGTGGGGGAAATTAAGTCTAT 1207
 QY 4058 TTTTCCCTACATGGATACACACATGTAATCAATCTTCACTGAAGGCCCTGCACTTC 4117
 DB 1208 TTTTCCCTACATGGATACACACATGTAATCAATCTTCACTGAAGGCCCTGCACTTC 1267
 QY 4118 TCTTAAACATAGTGTGTTTCTTTTCTTTTAAACAAAGTTAAGCTATGTTAATTA 4177
 DB 1268 TCTTAAACATAGTGTGTTTCTTTTCTTTTAAACAAAGTTAAGCTATGTTAATTA 1327
 QY 4178 AAAAATGCTGTCTGTCTCACTTCACTTGTGTTTAAAGCCATTCATTTGTGCTG 4237
 DB 1328 AAAAATGCTGTCTGTCTCACTTCACTTGTGTTTAAAGCCATTCATTTGTGCTG 1387
 QY 4238 TGTGTATTCATTAATCTTTTATACATTTCTGATGTGTAATTTGGTGTCTTGAAT 4297
 DB 1388 TGTGTATTCATTAATCTTTTATACATTTCTGATGTGTAATTTGGTGTCTTGAAT 1447
 QY 4298 ATCTTAATAAGATTCATTTGAATTAATTAATTAATTAATTAATTAATTAATTA 4357
 DB 1448 ATCTTAATAAGATTCATTTGAATTAATTAATTAATTAATTAATTAATTAATTA 1507
 QY 4358 AA 4359
 DB 1508 AA 1509

RESULT 8
 ADA55894
 ID ADA55894 standard; DNA; 1509 BP.

XX ADA55894;
 AC
 XX 20-NOV-2003 (first entry)

DE Gene encoding human secreted protein #73.

XX immunosuppressive; antiinflammatory; antiallergic;
 XX cytotoxic; cerebroprotective; neuroprotective; nootropic;
 KW cardiovascular; antiarteriosclerotic; gene therapy;
 KW human secreted protein; immune disorder; inflammation;
 KW respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;
 KW inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;
 KW multiple sclerosis; ischemic brain injury; Parkinson's disease;
 KW Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;
 KW triple helix formation; antisense gene therapy; forensic biology; ds;
 gene.
 XX Homo sapiens.
 OS

PN MO2002102994-A2.
XX
XX 27-DEC-2002.
XX
XX 19-MAR-2002; 2002MO-US008278.
XX
XX 21-MAR-2001; 2001US-0277340P.
XX 19-JUL-2001; 2001US-0306171P.
XX 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMAN-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI, 2003-167512/16.
XX P-PSDB; ADA56791.
XX
XX New human secreted polypeptides and polynucleotides, useful for
PT diagnosing, treating or preventing e.g. immune disorders, inflammatory
PT conditions, respiratory disorders, cancers, CNS disorders, or
PT neurodegenerative disorders.
XX
XX Claim 21; SEQ ID NO 83; 1754bp; English.
XX
XX The invention relates to 592 new human secreted polypeptides useful for
XX diagnosing, treating or preventing e.g. immune disorders, inflammatory
XX conditions, respiratory disorders, cancers, CNS disorders, or
XX neurodegenerative disorders, or polypeptides comprising an amino acid
XX sequence at least 95% identical to the new sequences. The polypeptides,
XX antibodies or antibody fragments that bind to the polypeptides, nucleic
XX acids encoding the polypeptides, agonists or antagonists that binds to
XX the polypeptide, are useful in preparing diagnostic or pharmaceutical
XX compositions for diagnosing, treating or preventing an e.g. immune
XX disorders, inflammatory conditions (e.g. inflammatory Bowel disease,
XX nephritis or Crohn's disease), respiratory disorders (e.g. asthma and
XX allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders
XX (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative
XX disorders (e.g. Parkinson's disease or Alzheimer's disease), and
XX cardiovascular disorders (e.g. atherosclerosis or myocarditis). The
XX polynucleotides are useful for chromosome identification, chromosome
XX mapping, for controlling gene expression through triple helix formation
XX or antisense DNA or RNA, in gene therapy, for identifying individuals
XX from minute biological samples, in forensic biology, and as hybridization
XX probes. The polypeptides are useful for as molecular weight markers on
XX sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
XX gels, to raise antibodies, for testing biological activities, and for
XX treating or preventing neural disorders, immune system disorders,
XX muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
XX renal, proliferative and/or cancerous diseases. This sequence corresponds
XX to a gene encoding one of the polypeptide of the invention. Note: The
XX sequence data for this patent did form part of the printed specification,
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1509 BP; 397 A; 272 C; 319 G; 521 T; 0 U; 0 Other;
SQ
Query Match 33.8%; Score 1478; DB 10; Length 1509;
Best Local Similarity 99.9%; Pred. No. 1.8e-292;
Matches 1509; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 2859 GATGTAACCTAATGAGCTTCCCATCTCTTGTAAATAATTTGTATGTGTACATCTT 2918
DB 9 GATGTAACCTAATGAGCTTCCCATCTCTTGTAAATAATTTGTATGTGTACATCTT 68
QY 2919 GGTCTCTCTCCCTCCGTTTGTAAATAATGAGATAGCACTCCAGGCACTTTGGTC 2978
DB 69 GGTCTCTCTCCCTCCGTTTGTAAATAATGAGATAGCACTCCAGGCACTTTGGTC 128
QY 2979 TCAGGTAGATCCCTATTAATCTGAAAGAAATAGAGCCAGACCTTGTGTCTCA 3038
DB 129 TCAGGTAGATCCCTATTAATCTGAAAGAAATAGAGCCAGACCTTGTGTCTCA 188
QY 3039 AATATATAGAAATGCTTTCTTTAGTCTTCAAGACTATTTGTGTGAAACAAAGTAGGGGT 3098

DB 189 AATATATAGAAATGCTTTCTTTAGTCTTCAAGACTATTTGTGTGAAACAAAGTAGGGGT 248
QY 3099 CTAATCTCTAAGAGTAGGGGCTTTTATCTTTAAAGAAATATGTCCCAATTATAG 3158
DB 249 CTAATCTCTAAGAGTAGGGGC-TTTATCTTTAAAGAAATATGTCCCAATTATAG 307
QY 3159 CACTTTAAGAGAAAGCAAGGTATGTA-GGGTGTGTGCTGGGCCATCAGTGGAGCAC 3217
DB 308 CACTTTAAGAGAAAGCAAGGTATGAGGGGTGTGTGCTGGGCCATCAGTGGAGCAC 367
QY 3218 GAAAGAGAAATGGATATACATTGTGGGAAGAGAAAGTAAGTTCTTCAGGGGCTCCCACT 3277
DB 368 GAAAGAGAAATGGATATACATTGTGGGAAGAGAAAGTAAGTTCTTCAGGGGCTCCCACT 427
QY 3278 GCTAAAGTTTTTTGTAGATGTGATCTGTGCTCTCGAATTTGACTTTAAAGAAATTA 3337
DB 428 GCTAAAGTTTTTTGTAGATGTGATCTGTGCTCTCGAATTTGACTTTAAAGAAATTA 487
QY 3338 TTCTGGCAGACATGATGATATCTTGGATGATCTTGTCTGTATTTCTCTTTGTGT 3397
DB 488 TTCTGGCAGACATGATGATATCTTGGATGATCTTGTCTGTATTTCTCTTTGTGT 547
QY 3398 GT 3457
DB 548 GT 607
QY 3458 GGGCTCTCTAATTAAGGAACTGTGTAACCTTCATTGCAAGAGATGTAGAGAAATA 3517
DB 608 GGGCTCTCTAATTAAGGAACTGTGTAACCTTCATTGCAAGAGATGTAGAGAAATA 667
QY 3518 GACTTAATTTCCACTGAGGCTCTCATCTCAACCTTAAGAGGAAATTTCTAAGAAAC 3577
DB 668 GACTTAATTTCCACTGAGGCTCTCATCTCAACCTTAAGAGGAAATTTCTAAGAAAC 727
QY 3578 TGGGCGAGATTTCTTTGTTCTCCATCTTTATGTGACAGGCTGTCAAGTTTCTTAC 3637
DB 728 TGGGCGAGATTTCTTTGTTCTCCATCTTTATGTGACAGGCTGTCAAGTTTCTTAC 787
QY 3638 TCTTACCTATGTGATATTTCTTGTGATAGTGTCAAAAAGAAAAGAACCCATCAGTGT 3697
DB 788 TCTTACCTATGTGATATTTCTTGTGATAGTGTCAAAAAGAAAAGAACCCATCAGTGT 847
QY 3698 CTTTGAACCTTGTCTTTGATCCCTGATTTCTTGTGATTTTACAGATGTGTGGTCTC 3757
DB 848 CTTTGAACCTTGTCTTTGATCCCTGATTTCTTGTGATTTTACAGATGTGTGGTCTC 907
QY 3758 TAATTTGGGTATGAGTTAGCAATTTAAACATTTGTTGTGSCCTTACCCAGGGAGCTC 3817
DB 908 TAATTTGGGTATGAGTTAGCAATTTAAACATTTGTTGTGSCCTTACCCAGGGAGCTC 967
QY 3818 CCCAGTTTCTGACTTGAAGTGAAGTGAAGAAATCCAGAGGTCTATCTGGCCAGATTT 3877
DB 968 CCCAGTTTCTGACTTGAAGTGAAGTGAAGAAATCCAGAGGTCTATCTGGCCAGATTT 1027
QY 3878 AAGTATATCTATTTCTTGTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3937
DB 1028 AAGTATATCTATTTCTTGTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1087
QY 3938 TCTTGAAGTTATTTCTCTGATTTGATCTTGTGATTTGATGAGAGGTTGGAGCTGATTT 3997
DB 1088 TCTTGAAGTTATTTCTCTGATTTGATCTTGTGATTTGATGAGAGGTTGGAGCTGATTT 1147
QY 3998 AGCAAGTTTCAAGTCAAAATTTACAGTGTGTGAGAGTGTGGGGGAAATTAATGCTTAT 4057
DB 1148 AGCAAGTTTCAAGTCAAAATTTACAGTGTGTGAGAGTGTGGGGGAAATTAATGCTTAT 1207
QY 4058 TTTTCCCTACATGAGATACAACTGTGATTTCAATCTTCACTGAAGCCCTGCACTTC 4117
DB 1208 TTTTCCCTACATGAGATACAACTGTGATTTCAATCTTCACTGAAGCCCTGCACTTC 1267
QY 4118 TCCATAAACAATATTTTGT 4177

Db 1268 TCCTAAACATAGTGTGTTTCTTTTACAAAGTTAAGCTAGTGAATTAATTA 1327
Qy 4178 AAAAAATGCTGTCTGTCTACTTCAAGCTTGTGTTTATGCCATTGATGTTGCTG 4237
Db 1328 AAAAAATGCTGTCTGTCTACTTCAAGCTTGTGTTTATGCCATTGATGTTGCTG 1387
Qy 4238 TGTGTATATCATTAATCTTTGATACCAATTTCTGATGTGTAATAATGTTGCTGTAAT 4297
Db 1388 TGTGTATATCATTAATCTTTGATACCAATTTCTGATGTGTAATAATGTTGCTGTAAT 1447
Qy 4298 ATCTTAAGAGAGTCAATGTAATTAATACTATGTCGCTTAAAAAATGTTGCTGTAAT 4357
Db 1448 ATCTTAAGAGAGTCAATGTAATTAATACTATGTCGCTTAAAAAATGTTGCTGTAAT 1507
Qy 4358 AA 4359
Db 1508 AA 1509
RESULT 9
ADQ86569
ID ADQ86569 standard; cDNA; 1755 BP.
XX
AC ADQ86569;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human tumour-associated antigenic target (TAT) cDNA sequence #3442.
XX
XX human; tumour-associated antigenic target; TAT; cytosolic; gene therapy;
KW cancer; cell proliferative disorder; gene; ss.
XX
XX Homo sapiens.
XX
PN MO2004060270-A2.
XX
PD 22-JUL-2004.
XX
PF 15-OCT-2003; 2003WO-US029126.
XX
PR 18-OCT-2002; 2002US-0418988P.
XX
PA (GETH) GENENTECH INC.
PA (WUTD/) WU T D.
PA (ZHOU/) ZHOU Y.
XX
PI Wu TD, Zhou Y;
XX
DR WPI; 2004-534300/51.
XX
XX New nucleic acid molecule and encoded polypeptide, for diagnosing,
PT preventing or treating cell proliferative disorders such as cancer.
XX
PS Claim 1; SEQ ID NO 3442; 5504pp; English.
XX
XX The present invention describes an isolated tumour-associated antigenic
CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a) -
CC (c). Also described: (1) an expression vector comprising the above
CC nucleic acid; (2) a host cell comprising the above expression vector; (3)
CC a process for producing a polypeptide; (4) an isolated polypeptide
CC comprising: (a) an amino acid sequence encoded by any of the above
CC nucleotide sequences; (b) an amino acid sequence encoded by the full-
CC length coding region of the above nucleotide sequences; or (c) a sequence
CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
CC comprising the above polypeptide fused to a heterologous polypeptide; (6)
CC an isolated antibody that binds to the above polypeptide; (7) a process
CC for producing the antibody; (8) an isolated oligopeptide that binds to
CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)
CC binding organic molecule that binds to the above polypeptide; (10) a
CC composition of matter comprising the above (chimeric) polypeptide;

CC antibody, oligopeptide or TAT binding organic molecule, in combination
CC with a carrier; (11) an article of manufacture comprising a container and
CC the composition of matter contained within the container; (12) methods of
CC inhibiting the growth of a cell that expresses the above protein, where
CC the growth of the cell is at least in part dependent upon a growth
CC potentiating effect of the above protein; (13) a method of
CC therapeutically treating a mammal having a cancerous tumour comprising
CC cells that express the above protein; (14) a method of determining the
CC presence of a protein in a sample suspected of containing the protein
CC described above; (15) methods of diagnosing the presence of a tumour in a
CC mammal; (16) a method for treating or preventing a cell proliferative
CC disorder associated with increased expression or activity of the above
CC protein; and (17) a method of binding an antibody, oligopeptide or
CC organic molecule to a cell that expresses the protein described above.
CC The TAT sequences have cytosolic activities, and can be used in gene
CC therapy. The composition and methods are useful for diagnosing,
CC preventing or treating cancer. The composition is also used for preparing
CC a medicament for the therapeutic treatment or diagnostic detection of a
CC cell proliferative disorder or cancer. The present sequence represents a
CC human TAT cDNA sequence from the present invention.
XX
SQ Sequence 1755 BP; 435 A; 502 C; 505 G; 313 T; 0 U; 0 Other;
Query Match 31.3%; Score 1370; DB 12; Length 1755;
Best Local Similarity 100.0%; Pred. No. 2.5e-270;
Matches 1370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 84 ATGAGATCCGCGGCTGTTCAAGAGGCGCGCCCTGCGCGGCATCTGAGCAGCGA 143
Db 1 ATGAGATCCGCGGCTGTTCAAGAGGCGCGCCCTGCGCGGCATCTGAGCAGCGA 60
Qy 144 CTGAGACTACGAGGCGGCTGCGGCTGCGGCGCGCGCGAGCAGCAGCTGAGGCGCGCGAA 203
Db 61 CTGAGACTACGAGGCGGCTGCGGCTGCGGCGCGCGAGCAGCAGCTGAGGCGCGCGAA 120
Qy 204 GCCACTTCTCCCGCGCACCCTTAAGCTCAAGAGCGCGGCGCGCGCTGCTCTCC 263
Db 121 GCCACTTCTCCCGCGCACCCTTAAGCTCAAGAGCGCGGCGCGCGCTGCTCTCC 180
Qy 264 CAGGCGGAGAGCGCGCGCTTCCCGCGCGCGCTGCGGCGCGCGCGCGCGCTTCTTA 323
Db 181 CAGGCGGAGAGCGCGCGCTTCCCGCGCGCGCTGCGGCGCGCGCGCGCGCTTCTTA 240
Qy 324 CTCGCGGCGGCTT 383
Db 241 CTCGCGGCGGCTT 300
Qy 384 CCTGTCCG 443
Db 301 CCTGTCCG 360
Qy 444 GCCGCGTGCAGCG 503
Db 361 GCCGCGTGCAGCG 420
Qy 504 GTGCG 563
Db 421 GTGCG 480
Qy 564 GCTGCGACGCGCTCGAGACCGCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
Db 481 GCTGCGACGCGCTCGAGACCGCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Qy 624 CTGCG 683
Db 541 CTGCG 600
Qy 684 AGGAAGAGCGCTTCTGCGGAGTGTGCGGCGCTGCGAGCGCTCGAGTGAAGAGTATG 743
Db 601 AGGAAGAGCGCTTCTGCGGAGTGTGCGGCGCTGCGAGCGCTCGAGTGAAGAGTATG 660
Qy 744 AAACAGATCTTGTGTCGATTTGAATTCATGAAACAGCAGCAGCAGCAGCAGCTGACGCGC 803

Db	661	AAACGATCTTCTGCTGCATTTGGAACTCATCCAAACAGCAGACGACGACTGCAGGCC	720
Qy	804	AAGAAAAAGGATTCGAGAGCTGAAAGTCAGAGAGAGACACGCTCTTGCTCGATTGAA	863
Db	721	AAGAAAAAGGATTCGAGAGACTGAAAGTCAGAGAGAGACACGCTCTTGCTCGATTGAA	780
Qy	864	CGTATGGAAGACCGGATGCGACTGGTTAAGAAAGGATAACGAGAAAGAAAGCACAGCTG	923
Db	781	CGTATGGAAGACCGGATGCGACTGGTTAAGAAAGGATAACGAGAAAGAAAGCACAGCTG	840
Qy	924	TTTCAAGGCTATGAAAATCTGAAGAGAGAGAGAAACAGAGCTATCTGAGAAAAATTAACTG	983
Db	841	TTTCAAGGCTATGAAAATCTGAAGAGAGAGAGAAACAGAGCTATCTGAGAAAAATTAACTG	900
Qy	984	GAGTGCACCGCGAGCTTTCCGAGACATCCGAGACTCTGCTCCAGACCCCTTCTCATGT	1043
Db	901	GAGTGCACCGCGAGCTTTCCGAGACATCCGAGACTCTGCTCCAGACCCCTTCTCATGT	960
Qy	1044	GGGGCGGATGGAAAAGGACATPAAAAAGAAATCCCATTTGGAAAGTACGAAAAGAAAGACT	1103
Db	961	GGGGCGGATGGAAAAGGACATPAAAAAGAAATCCCATTTGGAAAGTACGAAAAGAAAGACT	1020
Qy	1104	CCTGTTAAAAAGCTGGCTCCCTGAATTTTCAAAAAGTCAAAAACAAAACCTCTAAGCAGCT	1163
Db	1021	CCTGTTAAAAAGCTGGCTCCCTGAATTTTCAAAAAGTCAAAAACAAAACCTCTAAGCAGCT	1080
Qy	1164	CCTAATTAAAGAGAACCCCTGTGTGTTCTTATCTGAAACTGTTTGTAAACGTGAATTGAGG	1223
Db	1081	CCTAATTAAAGAGAACCCCTGTGTGTTCTTATCTGAAACTGTTTGTAAACGTGAATTGAGG	1140
Qy	1224	AGCCAAAGAAACCCCAAGAAAAGCCCCGGCTTTCAGTGGACACCCCAACCAAGACTTCCACT	1283
Db	1141	AGCCAAAGAAACCCCAAGAAAAGCCCCGGCTTTCAGTGGACACCCCAACCAAGACTTCCACT	1200
Qy	1284	CCCCAAAAAGGAGCCGAGACCCCATCCCAAGGAGAAAGCCTTCTCAAGTGAATGGAAGAT	1343
Db	1201	CCCCAAAAAGGAGCCGAGACCCCATCCCAAGGAGAAAGCCTTCTCAAGTGAATGGAAGAT	1260
Qy	1344	TTGCGCGTACCTTTCCACACACAGAAATGTATTGTGTCGTGGCACACGACTCCCCATCA	1403
Db	1261	TTGCGCGTACCTTTCCACACACAGAAATGTATTGTGTCGTGGCACACGACTCCCCATCA	1320
Qy	1404	CCGTTTACCATTACGGGAAATCCTCTCCAAAAGAGAGAGACTGTAGCAAG	1453
Db	1321	CCGTTTACCATTACGGGAAATCCTCTCCAAAAGAGAGAGACTGTAGCAAG	1370
RESULT 10			
ADQ84001			
ID	ADQ84001	standard; cDNA; 1755 BP.	
XX	ADQ84001;		
AC	ADQ84001;		
DT	07-OCT-2004	(first entry)	
XX			
DE	Human tumour-associated antigenic target (TAT) cDNA sequence #815.		
XX	human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;		
KM	cancer; cell proliferative disorder; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	W02004060270-A2.		
XX			
PD	22-JUL-2004.		
XX			
PF	15-OCT-2003; 2003WO-US029126.		
XX			
PR	18-OCT-2002; 2002US-0418988P.		
XX			
PA	(GETH) GENENTECH INC.		
PA	(WOTD)/ WU T D.		
PA	(ZHOU)/ ZHOU Y.		

Pt	New nucleic acid molecule and encoded polypeptide, for diagnosing,
Pt	preventing or treating cell proliferative disorders such as cancer.
Xx	
Xx	Claim 1; SEQ ID NO 815; 5504dp; English.
Ps	
Xx	
Xx	The present invention describes an isolated tumour-associated antigenic
Cc	target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
Cc	sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
Cc	(a); (c) the complement of (a) or (b); (d) a sequence that has 80%
Cc	sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-
Cc	(c). Also described: (1) an expression vector comprising the above
Cc	nucleic acid; (2) a host cell comprising the above expression vector; (3)
Cc	a process for producing a polypeptide; (4) an isolated polypeptide
Cc	comprising: (a) an amino acid sequence encoded by any of the above
Cc	nucleotide sequences; (b) an amino acid sequence encoded by the full-
Cc	length coding region of the above nucleotide sequences; or (c) a sequence
Cc	having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
Cc	comprising the above polypeptide fused to a heterologous polypeptide; (6)
Cc	an isolated antibody that binds to the above polypeptide; (7) a process
Cc	for producing the antibody; (8) an isolated oligopeptide that binds to
Cc	the above polypeptide; (9) a tumour-associated antigenic target (TAT)
Cc	binding organic molecule that binds to the above polypeptide; (10) a
Cc	composition of matter comprising the above (chimeric) polypeptide,
Cc	antibody, oligopeptide or TAT binding organic molecule, in combination
Cc	with a carrier; (11) an article of manufacture comprising a container and
Cc	the composition of matter contained within the container; (12) methods of
Cc	inhibiting the growth of a cell that expresses the above protein, where
Cc	the growth of the cell is at least in part dependent upon a growth
Cc	potentiating effect of the above protein; (13) a method of
Cc	therapeutically treating a mammal having a cancerous tumour comprising
Cc	cells that express the above protein; (14) a method of determining the
Cc	presence of a protein in a sample suspected of containing the protein
Cc	described above; (15) methods of diagnosing the presence of a tumour in a
Cc	mammal; (16) a method for treating or preventing a cell proliferative
Cc	disorder associated with increased expression or activity of the above
Cc	protein; and (17) a method of binding an antibody, oligopeptide or
Cc	organic molecule to a cell that expresses the protein described above.
Cc	The TAT sequences have cytostatic activities, and can be used in gene
Cc	therapy. The composition and methods are useful for diagnosing,
Cc	preventing or treating cancer. The composition is also used for preparing
Cc	a medicament for the therapeutic treatment or diagnostic detection of a
Cc	cell proliferative disorder or cancer. The present sequence represents a
Cc	human TAT cDNA sequence from the present invention.
Sq	
Sq	Sequence 1755 BP; 435 A; 502 C; 505 G; 313 T; 0 U; 0 Other;
Query Match	31.3%; Score 1370; DB 13; Length 1755;
Best Local Similarity	100.0%; Pred. No. 2.5e-270; Mismatches 0; Indels 0; Gaps 0;
Matches 1370; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Oy	84 ATGAGATCCGGCGGTTCACAGCGCGCGGCCCTGCCTGGCGGCAACTTGAGCAGCGA 143
Db	1 ATGAGATCCGGCGGTTCACAGCGCGGCCCTGCCTGGCGGCAACTTGAGCAGCGA 60
Oy	144 CTGAAGTAAGAGCGGGCTGCGGCGCTGCGGCCGAGACAAGACCTGGGCGGCCGAA 203
Db	61 CTGAAGTAAGAGCGGGCTGCGGCGCTGCGGCCGAGACAAGACCTGGGCGGCCGAA 120
Oy	204 GCCCACTTCCTCCCCCGGCAACCTGAAGTAAGAGCCGGGGCCCCCGCTGGCTCTCC 263
Db	121 GCCCACTTCCTCCCCCGGCAACCTGAAGTAAGAGCCGGGGCCCCCGCTGGCTCTCC 180
Oy	264 CAGGCGGAGACCCCGGCGCTTCCCGGCGCGGCTGCGGCGGCAAGGGCGGGGCTTTGTA 323
Db	181 CAGGCGGAGACCCCGGCGCTTCCCGGCGCGGCTGCGGCGGCAAGGGCGGGGCTTTGTA 240
Oy	324 CTCCTGGCGGGGCGGCCCTCCGGGACAGGAGAAAGACTTGGGTGGCTCTTG 383

Db	241	CTCCGCGCGGAGCGCCCGCGGACGACAGAAAGAGTGTGGGCGGTTTCGATGCCCTTG	30
QY	384	CCCTGTCCGCCCGCGGCACCAAGCAAGCGGCATTTGGGGGGAGCTCCGCGACGCGA	443
Db	301	CCCTGTCCGCCCGCGGCACCAAGCAAGCGGCATTTGGGGGGAGCTCCGCGACGCGA	360
QY	444	GCCGGCTCAGGCCCCCGGCCCAAGTATCAGGGGGGTGCTGCCATTCAGAGGGGCTCTC	503
Db	361	GCCGGCTCAGGCCCCCGGCCCAAGTATCAGGGGGGTGCTGCCATTCAGAGGGGCTCTC	420
QY	504	GTGGCGGGGCCAAAGAGCTTACGCGCTGGGCTGGGGACAAGGGTGGGGCGCTCCCC	563
Db	421	GTGGCGGGGCCCAAGAGCTTACGCGCTGGGCTGGGGACAAGGGTGGGGCGCTCCCC	480
QY	564	GCTGCGACCGCTTCGAGCCCGCGGGACCCCGACCACTACTCTGCGCGGCGCCGACCC	623
Db	481	GCTGCGACCGCTTCGAGCCCGCGGGACCCCGACCACTACTCTGCGCGGCGCCGACCC	540
QY	624	CTCGGGCCCAACCGCCACCGCGGGACCTCTGGCGGCTCAGGAGGGCAGATGAAAGATTG	683
Db	541	CTCGGGCCCAACCGCCACCGCGGGACCTCTGGCGGCTCAGGAGGGCAGATGAAAGATTG	600
QY	684	AGAAAGAGCCCTCTGGGGGGTGTGTGGCGGCTCGGGAGCTCCAGTCAAGCGCTGCGTC	743
Db	601	AGAAAGAGCCCTCTGGGGGGTGTGTGGCGGCTCGGGAGCTCCAGTCAAGCGCTGCGTC	660
QY	744	AAACGATCTTCTGTGCTCATTTGACCTCATCGAACAGCAGCAGCAGCGTGCAGGCG	803
Db	661	AAACGATCTTCTGTGCTCATTTGACCTCATCGAACAGCAGCAGCAGCGTGCAGGCG	720
QY	804	AAGGAAAAAGAGATGAGAGGCTGAAGTCAAGAGAGACACGCTCTTCTCTGGATTGAA	863
Db	721	AAGGAAAAAGAGATGAGAGGCTGAAGTCAAGAGAGACACGCTCTTCTCTGGATTGAA	780
QY	864	CGTATGAAAAGCGGATGAGCTGTGTAAAGAGGATTAAGAAAGAAAGGACAAAGCTG	923
Db	781	CGTATGAAAAGCGGATGAGCTGTGTAAAGAGGATTAAGAAAGAAAGGACAAAGCTG	840
QY	924	TTTTCAGGGCTATGAAACTGAAGAGAGAGAGAAACAGACTATCTGAGAAATTTAACTG	983
Db	841	TTTTCAGGGCTATGAAACTGAAGAGAGAGAGAAACAGACTATCTGAGAAATTTAACTG	900
QY	984	GAGTCCAGCGGAGCTTTCGAGACATCCAGACTCTGCTCCGCGGCTTCTCATGT	1044
Db	901	GAGTCCAGCGGAGCTTTCGAGACATCCAGACTCTGCTCCGCGGCTTCTCATGT	960
QY	1044	GGGCGGAGTGAAGAGCATTAAGAAATGCCCATTTGGAAAGTACAAAGAAAGAACT	1104
Db	961	GGGCGGAGTGAAGAGCATTAAGAAATGCCCATTTGGAAAGTACAAAGAAAGAACT	1022
QY	1104	CCTGTTAAAAAGCTGGCTCTGAATTTTCAAAAGTCAAAACAAAACTCTTAAGACTCT	1166
Db	1021	CCTGTTAAAAAGCTGGCTCTGAATTTTCAAAAGTCAAAACAAAACTCTTAAGACTCT	1088
QY	1164	CCTATTTAAAGAGAACCTGTGTTCTTATCTGAATCTGTTAACTGTTAACTGAATTGAG	1222
Db	1081	CCTATTTAAAGAGAACCTGTGTTCTTATCTGAATCTGTTAACTGTTAACTGAATTGAG	1144
QY	1224	AGCCAAAGAAACCCCAAGAAAGCCCCGGGTCTTCAATGSAACAACCCACACAGACTCTCACT	1288
Db	1141	AGCCAAAGAAACCCCAAGAAAGCCCCGGGTCTTCAATGSAACAACCCACACAGACTCTCACT	1206
QY	1284	CCCCAAAGGGAACCCAGCAACCATCCCAAGAGAAAGCTTCTCAAGTGAATAGAAAGAT	1344
Db	1201	CCCCAAAGGGAACCCAGCAACCATCCCAAGAGAAAGCTTCTCAAGTGAATAGAAAGAT	1266
QY	1344	TTGCGGTACTTTTCACACACAGAAATGTATTTGTGTGTTGGACACGAGCTCCGCCATCA	1404
Db	1261	TTGCGGTACTTTTCACACACAGAAATGTATTTGTGTGTTGGACACGAGCTCCGCCATCA	1322
QY	1404	CCGTTACCAATTACGGGAATCTCTCCAAAGAGAGAGABAATGTAGCAAG	1453
Db	1321	CCGTTACCAATTACGGGAATCTCTCCAAAGAGAGAGABAATGTAGCAAG	1370

RESULT 11
 ID ADO83399 standard; cDNA, 1755 BP.
 AC ADO83399;
 XX
 XX
 DT 07-OCT-2004 (first entry)
 XX
 XX Human tumour-associated antigenic target (TAT) cDNA sequence #213.
 DE human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
 KW cancer; cell proliferative disorder; gene; ss.
 XX
 OS Homo sapiens.
 PN WO2004060270-A2.
 PM
 PD 22-JUL-2004.
 XX
 XX 15-OCT-2003; 2003WO-US029126.
 PF 18-OCT-2002; 2002US-0418988P.
 PR
 PA (GETH) GENENTECH INC.
 PA (WUTD) WU T D.
 PA (ZHOU) ZHOU Y.
 XX
 PI Wu TD, Zhou Y;
 XX
 XX WPI, 2004-534300/51.
 PT New nucleic acid molecule and encoded polypeptide, for diagnosing,
 PT preventing or treating cell proliferative disorders such as cancer.
 XX
 XX Claim 1; SEQ ID NO 213; 5504pp; English.
 XX
 XX The present invention describes an isolated tumour-associated antigenic
 CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
 CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
 CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
 CC sequence identity to (a)-(c); or (e) a sequence that hybridizes to (a)-
 CC (c). Also described: (1) an expression vector comprising the above
 CC nucleic acid; (2) a host cell comprising the above expression vector; (3)
 CC a process for producing a polypeptide; (4) an isolated polypeptide
 CC comprising: (a) an amino acid sequence encoded by any of the above
 CC nucleotide sequences; (b) an amino acid sequence encoded by the full-
 CC length coding region of the above nucleotide sequences; or (c) a sequence
 CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
 CC comprising the above polypeptide fused to a heterologous polypeptide; (6)
 CC an isolated antibody that binds to the above polypeptide; (7) a process
 CC for producing the antibody; (8) an isolated oligopeptide that binds to
 CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)
 CC binding organic molecule that binds to the above polypeptide; (10) a
 CC composition of matter comprising the above (chimeric) polypeptide,
 CC antibody, oligopeptide or TAT binding organic molecule, in combination
 CC with a carrier; (11) an article of manufacture comprising a container and
 CC the composition of matter contained within the container; (12) methods of
 CC inhibiting the growth of a cell that expresses the above protein, where
 CC the growth of the cell is at least in part dependent upon a growth
 CC potentiating effect of the above protein; (13) a method of
 CC therapeutically treating a mammal having a cancerous tumour comprising
 CC cells that express the above protein; (14) a method of determining the
 CC presence of a protein in a sample suspected of containing the protein
 CC described above; (15) methods of diagnosing the presence of a tumour in a
 CC mammal; (16) a method for treating or preventing a cell proliferative
 CC disorder associated with increased expression or activity of the above
 CC protein; and (17) a method of binding an antibody, oligopeptide or
 CC organic molecule to a cell that expresses the protein described above.
 CC The TAT sequences have cytostatic activities and can be used in gene
 CC therapy. The composition and methods are useful for diagnosing,
 CC preventing or treating cancer. The composition is also used for preparing

CC a medication for the therapeutic treatment or diagnostic detection of a
CC cell proliferative disorder or cancer. The present sequence represents a
CC human YAT cDNA sequence from the present invention.

XX Sequence 1755 BP; 435 A; 502 C; 505 G; 313 T; 0 U; 0 Other;

Query Match 31.3%; Score 1370; DB 13; Length 1755;
Best Local Similarity 100.0%; Pred. No. 2.5e-270;
Matches 1370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 84 ATGAGATCCGCGGTGTTCAAGCGCGCGCCCTGCGCGGCAATCTGAGCAGCA 143
Db 1 ATGAGATCCGCGGTGTTCAAGCGCGCGCCCTGCGCGGCAATCTGAGCAGCA 60
QY 144 CTGAGACTACGAGCGGCTGCGCGCTGCGCGGCGCGGAGCAGAGCTTGGCGGAA 203
Db 61 CTGAGACTACGAGCGGCTGCGCGCTGCGCGGCGCGGAGCAGAGCTTGGCGGAA 120
QY 204 GCCCACTTCTCCCGCGGACCGTAACTCAAGAGCCGGGGGCCCGCTGCGCTCTCC 263
Db 121 GCCCACTTCTCCCGCGGACCGTAACTCAAGAGCCGGGGGCCCGCTGCGCTCTCC 180
QY 264 CAGGCGGAGAGCCCGCGCTTCCCGGCGCGCTGCGCGGCAAGGAGCGGGGCTTGT 323
Db 181 CAGGCGGAGAGCCCGCGCTTCCCGGCGCGCTGCGCGGCAAGGAGCGGGGCTTGT 240
QY 324 CTCCTCGCGCGGCGGCGCGCGCGGCGAGAGAAAGAGCTGCGGCTTGGTCCCTTG 383
Db 241 CTCCTCGCGCGGCGGCGCGCGCGGCGAGAGAAAGAGCTGCGGCTTGGTCCCTTG 300
QY 384 CCTGTGCGCGCGCGCGCGCGCGCGCGAGCAAGCGGCTTGGGGGGAGCTTGGCGAGCCGGA 443
Db 301 CCTGTGCGCGCGCGCGCGCGCGCGCGAGCAAGCGGCTTGGGGGGAGCTTGGCGAGCCGGA 360
QY 444 GCGCGCTGAGAGCCCGCGCGCGCGCGCGAGTATCAGGCGGTGCTGCAATTCAGAGCGGCTCTCTC 503
Db 361 GCGCGCTGAGAGCCCGCGCGCGCGCGCGAGTATCAGGCGGTGCTGCAATTCAGAGCGGCTCTCTC 420
QY 504 GTGCGCGCGCGCGCGCGCGCGCGCGAGTATCAGGCGGTGCGGAGCAAGGCTGCGGCTCTCTC 563
Db 421 GTGCGCGCGCGCGCGCGCGCGCGCGAGTATCAGGCGGTGCGGAGCAAGGCTGCGGCTCTCTC 480
QY 564 GCTGCGCAGCGGCTCTGGAACCGCGCGGAGCCCGCACACTACTTGTCCCGGCGCGCACCC 623
Db 481 GCTGCGCAGCGGCTCTGGAACCGCGCGGAGCCCGCACACTACTTGTCCCGGCGCGCACCC 540
QY 624 CTGCGCGCGCAGCGCAGCCGCGGAGCCCTGCGGAGCCCTCAAGTACAGGCGCGCTGCTTC 683
Db 541 CTGCGCGCGCAGCGCAGCCGCGGAGCCCTGCGGAGCCCTCAAGTACAGGCGCGCTGCTTC 600
QY 684 AGGAGAGGCGCTCTCGGCGGTGCTGCGGAGCCCTCAAGTACAGGCGCGCTGCTTC 743
Db 601 AGGAGAGGCGCTCTCGGCGGTGCTGCGGAGCCCTCAAGTACAGGCGCGCTGCTTC 660
QY 744 AAACAGATCTTCTGCTGCAATTGGAACCTCATGAAACAGCAGCAGCAGCTGCAAGCC 803
Db 661 AAACAGATCTTCTGCTGCAATTGGAACCTCATGAAACAGCAGCAGCAGCTGCAAGCC 720
QY 804 AAGGAAAGAGAGTCAAGAGCTGAGTCAAGAGAGCAGCTCTCTGCTCGAATTTGAA 863
Db 721 AAGGAAAGAGAGTCAAGAGCTGAGTCAAGAGAGCAGCTCTCTGCTCGAATTTGAA 780
QY 864 CGTATGAAAGCGAGTGAAGTGAATTAAGAAAGATTAAGAAAGCAAGCTG 923
Db 781 CGTATGAAAGCGAGTGAAGTGAATTAAGAAAGATTAAGAAAGCAAGCTG 840
QY 924 TTTCAGGCGCTATGAACTGAAGAGAGAGAAACAGAGCTATCTGAATAATTAAC 983
Db 841 TTTCAGGCGCTATGAACTGAAGAGAGAGAGAAACAGAGCTATCTGAATAATTAAC 900
QY 984 GAGTGCACGCGGAGCTTCCGAGACATCCCAACTTGTCTCCCAAGCCCTTCTCATGT 1043
Db 901 GAGTGCACGCGGAGCTTCCGAGACATCCCAACTTGTCTCCCAAGCCCTTCTCATGT 960
```

```
QY 1044 GGGCGGAGTGAAGAGGAGCATTAAGAAATCCCATTTTGAAGTACAGAAAGAAAGACT 1103
Db 961 GGGCGGAGTGAAGAGGAGCATTAAGAAATCCCATTTTGAAGTACAGAAAGAAAGACT 1020
QY 1104 CCTGTTAAAGAGTGGCTCCTGAATTTTCAAAAGTCAAAACAAAACCTCTAAGCACTCT 1163
Db 1021 CCTGTTAAAGAGTGGCTCCTGAATTTTCAAAAGTCAAAACAAAACCTCTAAGCACTCT 1080
QY 1164 CCTATTAAAGAGAAACCTGTGCTTCTTATCTGAATCTGTTTGAACGTGAATTTGAG 1223
Db 1081 CCTATTAAAGAGAAACCTGTGCTTCTTATCTGAATCTGTTTGAACGTGAATTTGAG 1140
QY 1224 AGCCAAAGAAACCCAGAAAGCCCGGTCTTCAAGTGAACACCCACAGACTCTCACT 1283
Db 1141 AGCCAAAGAAACCCAGAAAGCCCGGTCTTCAAGTGAACACCCACAGACTCTCACT 1200
QY 1284 CCCCAGAAAGGAGCCCGAGCCCATCCAGAGAGAAAGCTTCTCAAGTGAATGAAGAT 1343
Db 1201 CCCCAGAAAGGAGCCCGAGCCCATCCAGAGAGAAAGCTTCTCAAGTGAATGAAGAT 1260
QY 1344 TTGCGGTACCTTTCACACAGAAATGTAATTTGTGTGTGACAGCCGCTCCCATCA 1403
Db 1261 TTGCGGTACCTTTCACACAGAAATGTAATTTGTGTGTGACAGCCGCTCCCATCA 1320
QY 1404 CGGTTACATTAACGGAATCTCTCCAAAGAAAGAGAGACTGTACAG 1453
Db 1321 CGGTTACATTAACGGAATCTCTCCAAAGAAAGAGAGACTGTACAG 1370
```

RESULT 12

ADFS9738
ID ADFS9738 standard; cDNA; 2106 BP.

XX ADFS9738;

XX 12-FEB-2004 (first entry)

DE Human contig polynucleotide sequence SEQ ID NO:2105.

XX biological activity; genetic engineering; hybridisation probe; oligomer;

KW primer; chromosome mapping; gene mapping; recombinant protein production;

XX human; gene; ss.

XX Homo sapiens.

XX MO2003080795-A2.

PN 09-AUG-2002; 2002WC-US025485.

PF 09-AUG-2001; 2001US-0311261P.

PR (HYSE-) HYSEQ INC.

PA Tang YT, Yang Y, Wang Z, Wang G, Ma Y;

PI WPI; 2003-876918/81.

DR P-PsDB; ADF60130.

XX New polynucleotides, useful as hybridization probes, oligomers or

PT primers, for chromosome or gene mapping, for the recombinant production

PT of proteins, and for generating antisense DNA or RNA.

XX Example 2; SEQ ID NO 2105; 571bp; English.

XX The present invention describes isolated polynucleotide sequences (I),

CC which encode polypeptides (II) with biological activity. Also described:

CC (1) a vector comprising (I); (2) an expression vector comprising (I); (3)

CC a host cell genetically engineered to comprise (I) which is operatively

CC associated with a regulatory sequence that modulates expression of (I) in

CC the host cell; (4) a polypeptide (II) encoded by (I); (5) a composition

CC comprising the polypeptide of (4) and a carrier; (6) an antibody directed
CC against the polypeptide of (4); (7) detecting (1) or the polypeptide
CC of (4) in a sample; (8) identifying a compound that binds to the polypeptide
CC of (4); (9) producing the polypeptide of (4); and (10) a collection of
CC polynucleotides comprising at least one of the polynucleotide sequences
CC (1). The polynucleotides (1) can be used as hybridisation probes,
CC oligomers or primers, for chromosome or gene mapping, for the recombinant
CC production of proteins, and for generating antisense DNA or RNA. The
CC present sequence represents a human config polynucleotide sequence, which
CC is used in an example from the present invention.

SQ Sequence 2106 BP; 526 A; 569 C; 588 G; 422 T; 0 U; 1 Other;

Query Match	24.4%	Score 1066.6	DB 10	Length 2106
Best Local Similarity	95.0%	Pred. No. 3.5e-208		
Matches 1309, Conservative	0	Mismatches 40	Indels 29	Gaps 19

QY	105	GGGGCGGGGGCCCTGCGGGCGGCAATCTTGAGCAGCGACTGGACATACAGCGGGCTGGG	164
Db	2	GGGGCGGGGGCCCTGCGGGCGGCAATCTTGAGCAGCGACTGGACATACAGCGGGCTGGG	61
QY	165	GGCGTGGGCGGGCCCGA-GBACGAGCTTGGGGCGGCGGACCCACTTCTCCCGGCA	223
Db	62	GGCGTGGGCGGGCCCGATGAGAGACTTGGGGCGGCGGACCCACTTCTCCCGGCA	121
QY	224	CGGTAACTCAAGGAG-CGGGGGGCCCCGCTGGGCTCTCCCAAGGGC-GGGAGCCCCGGG	281
Db	122	CGTAAAGCTCAAGGAGCCCTGGGGCCCCGCTGGGCTCTCCCAAGGGCGGGAGCCCCGGG	181
QY	282	CCCTTCCCGGGCGGGCTGGGCGGGC---AAGGGCGGGGCTTGTAACTCCCGGCGGGGGC	337
Db	182	CCCTTCCCGGGCGGGCTGGGCGGGCGGGAGAGCGGGGGGCTTGTAACTCCCGGCGGGGGC	241
QY	338	GGCCCCGGGGCAGCAGAAAGAGAGCTGGGGCGGTTTCGTGCTTCCTTCCTGTCGCCCCC	397
Db	242	GGCCCCGGGGCAGCAGAAAGAGAGCTGGGGCGGTTTCGTGCTTCCTTCCTGTCGCCCCC	301
QY	398	GGCCACCAAGCAAGCGGGCATTTGGGGGGGGAGGCTGCGGAGCGGAGCGGGCGTGCAGGCC	457
Db	302	GGCCACCAAGCAAGCGGGCATTTGGGGGGGGAGGCTGCGGAGCGGAGCGGGCGTGCAGGCC	361
QY	458	CGGGCCCAAGATTCAGGGGGTGTGCTGCCATTTGAGCGGGCTCTTGTGAGCGGGCGCAA	517
Db	362	CGGGCCCAAGATTCAGGGGGTGTGCTGCCATTTGAGCGGGCTCTTGTGAGCGGGCGCAA	421
QY	518	AGAAGCTTACGCCCTGGGCTGGGGGACAAAGGTGGGGGGGCTTCCCGCTGCCACGGCTTC	577
Db	422	AGAAGCTTACGCCCTGGGCTGGGGGACAAAGGTGGGGGGGCTTCCCGCTGCCACGGCTTC	481
QY	578	GGACCCCGGGGGACCCCGACCACTACCTTGGCCGGGGCGGCACCCCTGGCGCCACCGG	637
Db	482	GGACCCCGGGGGACCCCGACCACTACCTTGGCCGGGGCGGCACCCCTGGCGCCACCGG	541
QY	638	CACCGCCGGGACCTTGGCGGGCCAGCGAGGGCAGATGGAAGATATGAGGAAGAGCCCTCT	697
Db	542	CACCGCCGGGACCTTGGCGGGCCAGCGAGGGCAGATGGAAGATATGAGGAAGAGCCCTCT	601
QY	698	CGGGGGGT-GGTGGCGGGCTTGGGAGGCTCCAGTTCAGGCGCGCTGCTCAAAACAGTCTTTC	756
Db	602	CGGGGGGTGGGTGGCGGGCTTGGGAGGCTCCAGTTCAGGCGCGCTGCTCAAAACAGTCTTTC	661
QY	757	TGCTCAATTGGACCTCATTCGAACAGCAGCAGAGAGCTGACGAGGCGCAAGGAAAGAGAG	816
Db	662	TGCTCAATTGGACCTCATTCGAACAGCAGCAGAGAGAGCTGACGAGGCGCAAGGAAAGAGAG	721
QY	817	TCGAGAGCTGAAGTTCAGAGAGAGACAGCTCTCTTGTCTC-GGATTAACGTATGGAAGG	875
Db	722	TCGAGAGACTGAAGTTCAGAGAGAGACAGCTCTCTTGTCTC-GGATTAACGTATGGAAGG	781
QY	876	CGGATGCA-GCTGGTAAAGAAAGATTAAGAGAAAGAAAGGCAACACTGTTTTCAGGGCTA	934
Db	782	CGGATGCAAGCTGGTAAAGAAAGATTAAGAGAAAGAAAGGCAACACTGTTTTCAGGGCTA	841

OY	935	TGAACGTGAAGAGAGAGA -GGAACACAGCATCTGTGAAAATATPAACTGAGGCCAGC	993
Dd	842	TGAACGTGAAGAGAGAGAGAGAACCCGAGCTATCTGGAAAATAATPAACTGAGGCCAGC	901
OY	994	CGAGCTTTCGAGACA ---TCCAGACTCTGCCTCCAGCCCTTC -TGATGTGGCGG	1049
Dd	902	CCGGAGCTTTCGAGACATCCCAAGACTGTGCCTCCAGCCCTTCTTATGTGGGCGG	961
OY	1050	AGTGG -AAGGGACATPAAAGGAATCCCAATT -GGAAGTACAGAAAAGAAAGA ---CTC	1104
Dd	962	AGTGAAGAAAGGACATPAAAGGAATCCCAATTGGAGATACAGAAAAGAAAGACTCC	1021
OY	1105	CTGTAAAGAGCTGGC -TCTGTAATTTTCAAAGTCAAACAAACAAACCTCTAGACACTT	1163
Dd	1022	CTGTAAAGAGCTGGCTCTCTGAAATTTTCAAAGTCAAACAAACAAACCTCTAGACACTT	1081
OY	1164	CCATAT -AAAAGAGAACCTGTGGTTCCTT -ATCGAAACTGTTTGTAAACGTGAATT	1219
Dd	1082	CTATATTTAAAGAGGAACCTGTGGTTCCTTATCTTGAACCTTTTGTAAACGTGAATT	1141
OY	1220	GA -GAGCCCAAGAAACCCCAAGAAAAGCCCGGTCCTT -CAGTGACACCCCAACCAAGACT	1276
Dd	1142	GAGGGAGCCCAAGAAACCCCAAGAAAAGCCCGGTCCTTCAGTGAGACACCCCAACCAAGACT	1201
OY	1277	CTCCACTTCCCAAAAGGAGACCCAGACCCATCCCAAGAGAGAAAGCTTTCTCAAGTGAAT	1336
Dd	1202	CTCCACTTCCCAAAAGGAGACCCAGACCCCAATCCCAAGAGAGAAAGCTTTCTCAAGTGAAT	1261
OY	1337	AGAAGATTTTGCCTGACTTTTCACCAACAGAAATGATTTTGTGTGTGGACACAGCC -TC	1395
Dd	1262	AGAAGATTTTGCCTGACTTTTCACCAACAGAAATGATTTTGTGTGTGGACACAGCCCTTC	1321
OY	1396	CCCATATCACGGTTACATTTACGGGATATCTCTCAAAGAAAGAGAGACGTATAGCAAG	1453
Dd	1322	CCCATATCACGGTTACATTTACGGGATATCTCTCAAAGAAAGAGAGACGTATAGCAAG	1379

RESULT 13

ID	AAFL
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	1
30	1
31	1
32	1
33	1
34	1
35	1
36	1
37	1
38	1
39	1
40	1
41	1
42	1
43	1
44	1
45	1
46	1
47	1
48	1
49	1
50	1
51	1
52	1
53	1
54	1
55	1
56	1
57	1
58	1
59	1
60	1
61	1
62	1
63	1
64	1
65	1
66	1
67	1
68	1
69	1
70	1
71	1
72	1
73	1
74	1
75	1
76	1
77	1
78	1
79	1
80	1
81	1
82	1
83	1
84	1
85	1
86	1
87	1
88	1
89	1
90	1
91	1
92	1
93	1
94	1
95	1
96	1
97	1
98	1
99	1
100	1

81.93 standard; DNA; 1529 BP.

AC AAF1

DT 14-M

DE Lung

KW Huma

KW gast

KW prod

Homo OS

PN WO20

PD 21-S

PF 08-M

PR 12-M

PA (HUM)

XXI

XX
2364

DR P-PS

Lung PT

PT such as lung cancer.
XX
PS Claim 1, Page 676-677; 1425bp; English.
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytosolic; cardiovascular;
CC immunomodulatory; muscular active general; vulnary; gastrointestinal
CC general; nephrologic; antileptive; gynecologic; or antibacterial
CC activity. The invention also includes antibodies specific for the protein
CC or polynucleotide sequences. The lung cancer associated polynucleotide
CC sequences may be used for detection of lung cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The proteins may be used to treat disorders such as
CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
CC cardiovascular, renal, and proliferative disorders. The proteins may also
CC be used in the treatment of wounds and infectious diseases.
CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
CC used in the course of the invention for the identification and
CC characterization of the polynucleotide and protein sequences
XX
SQ Sequence 1529 BP; 460 A; 344 C; 379 G; 344 T; 0 U; 2 Other;
Query Match 19.1%; Score 837; DB 3; Length 1529;
Best Local Similarity 100.0%; Pred. No. 3.3e-161;
Matches 837; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 617 GCCACCCCTTCCGCGCCCAACCCGCGGACCTTGGCGCGCAAGCGGCGAGATTGAA 676
DB 2 GCCACCCCTTCCGCGCCCAACCCGCGGACCTTGGCGCGCAAGCGGCGAGATTGAA 61
QY 677 GAGTATGAGAAAGCCCTTCCGCGGAGTGTGCGGCTCGGAGACCTTCAGTACGCGC 726
DB 62 GAGTATGAGAAAGCCCTTCCGCGGAGTGTGCGGCTCGGAGACCTTCAGTACGCGC 121
QY 737 CTGCGCTCAACAGATCCTTCTGCTGCAATTGACCTTATGAAACAGAGAGAGACGACT 726
DB 122 CTGCGCTCAACAGATCCTTCTGCTGCAATTGACCTTATGAAACAGAGAGAGACGACT 181
QY 797 GCAGGCCCAAGAAAGAGATCGAGAGCTGAAGTCAAGAGAGACAGCTCTTGTCTCG 856
DB 182 GCAGGCCCAAGAAAGAGATCGAGAGCTGAAGTCAAGAGAGACAGCTCTTGTCTCG 241
QY 857 GATTGAACTGATGAAAGGCGGATGCGCTGTAAAGAGATTAACGAGAAAGAGCA 916
DB 242 GATTGAACTGATGAAAGGCGGATGCGCTGTAAAGAGATTAACGAGAAAGAGCA 301
QY 917 CAAGCTGTTCAAGGCGTATGAAATGAAAGAGAGAGAAACAGAGTATCTGAGAAAT 976
DB 302 CAAGCTGTTCAAGGCGTATGAAATGAAAGAGAGAGAAACAGAGTATCTGAGAAAT 361
QY 977 TAAACTGAGATGCGAGCGGAGCTTTCGAGACATCCGACACTCTGCTCCCAAGCCTT 1036
DB 362 TAAACTGAGATGCGAGCGGAGCTTTCGAGACATCCGACACTCTGCTCCCAAGCCTT 421
QY 1037 CTGATGCGGCGGAGTGAAGAGCAATTAAGAAATCCCAATTGGAAGTACAGAAAG 1096
DB 422 CTGATGCGGCGGAGTGAAGAGCAATTAAGAAATCCCAATTGGAAGTACAGAAAG 481
QY 1097 AAAGACTCCTGTTAAAGAGTGTGCTCTGAAATTTCAAAAGTCAAAACAAAACTCCTAA 1156
DB 482 AAAGACTCCTGTTAAAGAGTGTGCTCTGAAATTTCAAAAGTCAAAACAAAACTCCTAA 541
QY 1157 GCACCTCTCTATTAAAGAGAACCTGTGGTCTCTTATCTGAACCTGTTTGTAAAGTGA 1216
DB 542 GCACCTCTCTATTAAAGAGAACCTGTGGTCTCTTATCTGAACCTGTTTGTAAAGTGA 601
QY 1217 ATTGAGAGCCAAAGAAACCCCAAGAAAGCCCGGTCTTCAAGTGAACCCCAACAGACT 1276
DB 602 ATTGAGAGCCAAAGAAACCCCAAGAAAGCCCGGTCTTCAAGTGAACCCCAACAGACT 661
QY 1277 CTCACACTCCCAAAAGGAGACCAGACCCATCCCAAGAGAAAGCCTTCTCAAGTGAAT 1336

DB 662 CTCACACTCCCAAAAGGAGACCAGACCCATCCCAAGAGAAAGCCTTCTCAAGTGAAT 721
QY 1337 AGAAGATTTGCCGATACCTTTCCACCAAGAAATGATTTGTGCTGTGACACAGCCTCC 1396
DB 722 AGAAGATTTGCCGATACCTTTCCACCAAGAAATGATTTGTGCTGTGACACAGCCTCC 781
QY 1397 CCATATACCGTTTACCTTTACGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG 1453
DB 782 CCATATACCGTTTACCTTTACGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG 838
RESULT 14
AACT6874
ID AACT6874 standard; cDNA; 2887 BP.
XX
AC AACT6874;
XX
DT 08-FEB-2001 (first entry)
DE Human ORF2429 polynucleotide sequence SEQ ID NO:4857.
XX
KW Human; open reading frame; ORF; detection; cytosolic; hepatotropic;
KW vulnary; antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiac;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatologic; immunosuppressive; antineoplastic;
KW antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;
KW antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypochloridism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN MO20058473-A2.
XX
PD 05-OCT-2000.
XX
PR 31-MAR-2000; 2000MO-US008621.
XX
PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkete RA, Leach M;
XX
DR WPI; 2000-602362/57.
XX
DR P-PDB; AAB42665.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 5, Page 4044-4045; 5507bp; English.
XX
CC AACT4446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORF2 open reading frames 1 to 3161. The ORF2
CC sequences have activities such as: cytosolic; hepatotropic; vulnary;
CC antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
CC anticonvulsant; antidiabetic; immunosuppressant; immunostimulant;
CC cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatologic; immunosuppressive; antineoplastic; antihypertensive;
CC antiviral; antifungal; antineoplastic; antihypertensive; antineoplastic. The
CC sequences can be used for determining the presence of or predisposition
to, or preventing or treating pathological conditions associated with an

ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive

Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 0 U; 2 Other;

Query Match 18.0%; Score 785.4; DB 3; Length 2887;
Best Local Similarity 99.7%; Pred. No.1.5e-150;
Matches 797; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
QY 655 CGGCACGCGGCGAGATGGAAGATGAGAAAGACCCCTCTCGGGGGTGTGGCGGCT 714
DB 3 CGGCCGCGGCGAGATGGAAGATGAGAAAGACCCCTCTCGGGGGTGTGGCGGCT 62
QY 715 CGGAGCGCTCCAGTCAGGCGCGCTCTCAACAGATCTTCTGCAATTGCACTCA 774
DB 63 CGGAG-CCTCAGTCAGGCGCGCTCTCAACAGATCTTCTGCAATTGCACTCA 121
QY 775 TCGAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 834
DB 122 TCGAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 101
QY 835 AGAGAGACAGCGCTCTCTGCTCGATGTAAGATGAAAGCGGAGTCACTGGTAAAG 894
DB 182 AGAGAGACAGCGCTCTCTGCTCGATGTAAGATGAAAGCGGAGTCACTGGTAAAG 241
QY 895 AGGATACGAGAAAGAAAGCAAGCTGTTTCAAGGCTTGAACCTGAAGAGAGAG 954
DB 242 AGGATACGAGAAAGAAAGCAAGCTGTTTCAAGGCTTGAACCTGAAGAGAGAG 301
QY 955 AAACAGAGCTATCTGAGAAATTAATGAGTGCAGCGGAGCTTCCGAGCATCCC 1014
DB 302 AAACAGAGCTATCTGAGAAATTAATGAGTGCAGCGGAGCTTCCGAGCATCCC 361
QY 1015 AGACTGCTCCCAAGCCCTTCTCATGTGGCGGAGTGAAGAGGACATTAAGAAAT 1074
DB 362 AGACTGCTCCCAAGCCCTTCTCATGTGGCGGAGTGAAGAGGACATTAAGAAAT 421
QY 1075 CCCCATTGGAAGTACGAAAGAAAGACTCTGTTAAAGAGCTGCTCGAATTTCAA 1134
DB 422 CCCCATTGGAAGTACGAAAGAAAGACTCTGTTAAAGAGCTGCTCGAATTTCAA 481
QY 1135 AAGTCAAAAGAAAGCTCTAAGACCTCTAATTAAGAGGAAAGCTGAGTTCTTAT 1194
DB 482 AAGTCAAAAGAAAGCTCTAAGACCTCTAATTAAGAGGAAAGCTGAGTTCTTAT 541
QY 1195 CTGAAACTGTTTGTAAAGTGAATTTGAGAGCCAAAGAACCCCGAGAGCTT 1254
DB 542 CTGAAACTGTTTGTAAAGTGAATTTGAGAGCCAAAGAACCCCGAGAGCTT 601
QY 1255 CAGTGAACACCCCAAGAACTCTCCATCCCAAAAGGAGCCAGACCCATCCCAAG 1314
DB 602 CAGTGAACACCCCAAGAACTCTCCATCCCAAAAGGAGCCAGACCCATCCCAAG 661
QY 1315 AGAAGGCTTCTCAAGTGAATGAAATTTGCCGTACTTTCCACAAGAAATGAT 1374
DB 662 AGAAGGCTTCTCAAGTGAATGAAATTTGCCGTACTTTCCACAAGAAATGAT 721
QY 1375 TGTGCTTGTGCAACAGCTCTCCCATACCGTTACCTTACCGGAATCTCTCAAGA 1434
DB 722 TGTGCTTGTGCAACAGCTCTCCCATACCGTTACCTTACCGGAATCTCTCAAGA 781
QY 1435 AGGAGAGACTGTAGCAAG 1453
DB 782 AGGAGAGACTGTAGCAAG 800
```

RESULT 15
ADFS7688
ID ADFS7688 standard; cDNA; 1143 BP.

XX ADFS7688;

AC ADFS7688;

XX 12-FEB-2004 (first entry)

DE Human polynucleotide sequence SEQ ID NO:55.

XX biological activity; genetic engineering; hybridisation probe; oligomer;

KW primer; chromosome mapping; gene mapping; recombinant protein production;

XX human; gene; ss.

OS Homo sapiens.

PN WO2003080795-A2.

PD 02-OCT-2003.

XX 09-AUG-2002; 2002WO-US025485.

PR 09-AUG-2001; 2001US-0311261P.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Yang Y, Wang Z, Weng G, Ma Y;

XX WPI; 2003-876918/81.

DR P-PsDB; ADFS8688.

XX New polynucleotides, useful as hybridization probes, oligomers or

PT primers, for chromosome or gene mapping, for the recombinant production

PT of proteins, and for generating antisense DNA or RNA.

PS Claim 1; SEQ ID NO 55; 571bp; English.

XX The present sequence represents an isolated polynucleotide sequence (I)

CC from the present invention, which encodes a polypeptide (II) with

CC biological activity. Also described: (1) a vector comprising (II); (2) an

CC expression vector comprising (I); (3) a host cell genetically engineered

CC to comprise (I) which is operatively associated with a regulatory

CC sequence that modulates expression of (I) in the host cell; (4) a

CC polypeptide (II) encoded by (I); (5) a composition comprising the

CC polypeptide of (4) and a carrier; (6) an antibody directed against the

CC sample; (8) identifying a compound that binds to the polypeptide of (4);

CC (9) producing the polypeptide of (4); and (10) a collection of

CC polynucleotides comprising at least one of the polynucleotide sequences

CC (I). The polynucleotides (I) can be used as hybridisation probes,

CC oligomers or primers, for chromosome or gene mapping, for the recombinant

XX production of proteins, and for generating antisense DNA or RNA.

XX Sequence 1143 BP; 344 A; 287 C; 260 G; 252 T; 0 U; 0 Other;

Query Match 17.1%; Score 748.2; DB 10; Length 1143;
Best Local Similarity 94.1%; Pred. No.4.6e-143;
Matches 808; Conservative 0; Mismatches 3; Indels 48; Gaps 1;

```
QY 1064 TAAAGAAATCCCACTTTGGAAGTACAGAAAGAAAGACTCTGTTAAAGAGTGCCTC 1123
DB 333 TACGTGAATTCCTCCATTTGGAAGTACAGAAAGAAAGACTCTGTTAAAGAGTGCCTC 392
QY 1124 TGAATTTTCAAAAGTCAAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCA 1183
DB 393 TGAATTTTCAAAAGTCAAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCA 452
QY 1184 TGTGCTTCTTATCGAAGCTGTTTGTAAAGTGAATTTGAGAGGCAAGAAAGTCAAG 1243
DB 453 TGTGCTTCTTATCGAAGCTGTTTGTAAAGTGAATTTGAGAGGCAAGAAAGTCAAG 512
QY 1244 GCCCGGCTTTCAGTGAACACCCCAAGAACTCTCCATCCCAAAAGGAGCCAGCAC 1303
```

```
Db      ||||||| 513 GCCCGGTCTTGAGTGAACCCACCAAGACTCTCACTCCCAAAAGGACCCAGCAC 572
QY      ||||||| 1304 CCATCCCAAGAGAAAGCCTTCTCAAGTAGATAGAAGATTTGCCGTACCTTTCCACAC 1363
Db      ||||||| 573 CCATCCCAAGAGAAAGCCTTCTCAAGTAGATAGAAGATTTGCCGTACCTTTCCACAC 632
QY      ||||||| 1364 AGAAATGTATTTGTGTGTTGGACACGCTCCCCCATCAAGTTTCCATTACGGGAATC 1423
Db      ||||||| 633 AGAAATGTATTTGTGTGTTGGACACGCTCCCCCATCAAGTTTCCATTACGGGAATC 692
QY      ||||||| 1424 CTCTCCAAAGAGAGAGAGACTGTAGCAAGTGTCTGATGCCATCAAGTTGACAGAGA 1483
Db      ||||||| 693 CTCTCCAAAGAGAGAGAGACTGTAGCAAGTGTCTGATGCCATCAAGTTGACAGAGA 721
QY      ||||||| 1484 AACTTCAAGTCTTGAGCTGTTCTTCTTGAGAGGACCACTCACTGAGAGCTTTAAGGACCC 1543
Db      ||||||| 722 -----TTCCCTTCTTGAGAGGACCACTCACTGAGAGCTTTAAGGACCC 764
QY      ||||||| 1544 AAATCCTCAGACCTTTTGAGAAAGCTGAGTGAAGTGTGTTTCGAAGCGGATGCAAA 1603
Db      ||||||| 765 AAATCCTCAGACCTTTTGAGAAAGCTGAGTGAAGTGTGTTTCGAAGCGGATGCAAA 824
QY      ||||||| 1604 ACTGAGCTGATGAGAGAGAGAAAGATGAGATATTCAAGAGATCAGGAGACAAAG 1663
Db      ||||||| 825 ACTGAGCTGATGAGAGAGAGAGAAAGATGAGATATTCAAGAGATCAGGAGACAAAG 884
QY      ||||||| 1664 AATTTACAGCGACTGCACTCAAAATGTATTAAGAAAGAAATTCAGGAATCTGAGCC 1723
Db      ||||||| 885 AATTTACAGCGACTGCACTCAAAATGTATTAAGAAAGAAATTCAGGAATCTGAGCC 944
QY      ||||||| 1724 TGAAGTTACCTCATTTTCCCTGAGCAGATGATGTTGAAAGTTGATGATTAACCCCTT 1783
Db      ||||||| 945 TGAAGTTACCTCATTTTCCCTGAGCAGATGATGTTGAAAGTTGATGATTAACCCCTT 1004
QY      ||||||| 1784 CTGCTGTGTGATGATTTGAGACGACCATTAACCAAAATTAACCAAGATTTGAGCT 1843
Db      ||||||| 1005 CTGCTGTGTGATGATTTGAGACGACCATTAACCAAAATTAACCAAGATTTGAGCT 1064
QY      ||||||| 1844 ACCCTGTGTGATGAGCGTATGCCGATGCAAGTTGAGATCCAGAGAGCAAAACACCTCA 1903
Db      ||||||| 1065 ACCCTGTGTGATGAGCGTATGCCGATGCAAGTTGAGATCCAGAGAGCAAAACACCTCA 1124
QY      ||||||| 1904 CCGGACGTGAGGAATAAG 1922
Db      ||||||| 1125 CCGGACGTGAGGAATAAG 1143
```

Search completed: March 24, 2005, 11:38:25
Job time : 2151 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2005, 14:51:43 ; Search time 13096 Seconds
(without alignments)
12707.467 Million cell updates/sec

Title: US-10-054-935-1

Perfect score: 4372
Sequence: 1 cagtcctcgaccccccgcac.....aaaaaaaaaaaaaaaa 4372

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 1903234700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_est8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3691	84.4	3785	3	CR749360 Homo sapi
2	1950.4	44.6	3161	3	AK014463 Mus muscu
3	1798.6	41.1	2839	3	AK053719 Mus muscu
4	980.8	22.4	1304	3	AK015496 Mus muscu
5	897	20.5	1096	4	BM547318 AGENCOURT
6	805.6	18.4	896	5	BX329236 BX329236
7	805.6	18.4	927	5	BH857520 AGENCOURT
8	773.6	17.7	888	4	BH517114 AGENCOURT
9	770.8	17.6	817	4	BI917709 601183689
10	766	17.5	792	5	BH929398 AGENCOURT
11	758.6	17.4	769	5	CN355147 17004708
12	756.2	17.3	765	5	BQ014717 UI-H-BD1-
13	745.8	17.1	1032	4	BM475963 AGENCOURT
14	744.6	17.0	891	5	BH898842 AGENCOURT
15	737.4	16.9	931	5	BQ948768 AGENCOURT
16	734.2	16.8	773	5	BH612708 UI-M-FR0-
17	725.6	16.6	870	5	BQ229745 AGENCOURT
18	714	16.3	715	5	BM981391 UI-CF-BN1
19	713	16.3	921	5	BM160445 AGENCOURT
20	711	16.3	729	5	BH004676 UI-H-R10-
21	705	16.1	743	7	CN355143 17006001
22	703	16.1	712	6	CA447821 UI-H-R10-
23	701.8	16.1	902	5	BH517052 AGENCOURT
24	701.2	16.0	769	7	CK635536 UI-M-HN0-

C 25	697	15.9	697	5	BU682595	UI-CF-BE1
C 26	690.8	15.8	753	6	CA313331	UI-CF-FN0
C 27	689	15.8	704	7	CN355144	170005326
C 28	685	15.7	697	4	BI820795	603034002
C 29	679.8	15.5	864	4	BG289135	60287523
C 30	678.4	15.5	762	7	CK471650	AGENCOURT
C 31	678.2	15.5	745	2	BE907276	601500975
C 32	663.8	15.2	757	7	CV109405	AGENCOURT
C 33	663.2	15.2	933	5	BQ876347	AGENCOURT
C 34	662.2	15.1	783	7	CK478106	AGENCOURT
C 35	658.4	15.1	788	2	BE792962	601585622
C 36	656.4	15.0	716	6	CD639095	AGENCOURT
C 37	654.4	15.0	951	4	BG178033	602327595
C 38	652	14.9	756	6	CB851322	UI-CF-BE1
C 39	651.8	14.9	704	7	CF531438	UI-M-FY0-
C 40	648.6	14.8	854	4	BI665707	603287711
C 41	648.6	14.8	950	4	BG386583	602454831
C 42	648.2	14.8	686	5	BM970469	UI-CF-BE1
C 43	642.4	14.7	644	5	BX952313	DKFZP781M
C 44	640.6	14.7	700	1	AU252433	AU252433
C 45	640.4	14.6	643	6	CA447820	UI-H-R10-

ALIGNMENTS

RESULT 1
CR749360
LOCUS CR749360
DEFINITION Homo sapiens mRNA, cDNA DKFZp686J17211 (from clone DKFZp686J17211).
ACCESSION CR749360
VERSION CR749360.1 GI:51476445
KEYWORDS HTC.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Well, B., Auld, C., Osanger, A., Robo, G., Han, M. and Wiemann, S.
CONSTRM The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de; sequenced by BWPZ (Biomedical Research Center at the Heinrich-Heine-University, Dueseldorf/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp686J17211) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering: http://www.rzpd.de/cgi-bin/products/c1.cgi?cloneID=DKFZp686J17211 Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.
location/Qualifiers
1. 3785
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686J17211"
/tissue_type="rectum tumor"
/clone_lib="686 (synonym: hlc3). Vector psport1_sfi; host DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
/note="hypothetical protein, differentially spliced"
1. 3785
/gene="DKFZp686J17211"
107. 1348
/gene="DKFZp686J17211"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH18213.1"

/db_xref="GI:51475446"
translation="MRKSPUGGGGSGASSQAACLKQILLQLDLIEQQQQLQAKK
ELBELSERDITLLARIERMERRQVLVKONEKERHLFGYETEEREELSKILE
COPELSETSQTLPRPSPSCGRSGKHRSKPSFGSTKTPVKLAEPSVKTKPKH
SPKEBPCGSLSTPCKRELRSQETPEKPSVDTPPLSTPKGKSTHKEKAFSS
IEDLPYLPTTBWYLCRMHOPRSLPLRSPSKKEKRPSCMPSSVAGTSVLAPDS
MRDHSVEPLRDPMPSPDLBNDLSVFSKHAKLEBKRKRKRDRIREORLQQLQ
LMRYKKGIQESPEVTSFPEPDDVESLMTTFPLVVAFRPLPLTTPQNFELPWL
EESRRLLEIQRKQTPHRTCRK"

ORIGIN

Query Match 84.4%; Score 3691; DB 3; Length 3785;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 3700; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 645 GGGACCCCTGGCGCCACGAGGCGAGATGAGAGATGAGAGAGCCCTCGGGGT 704
DB 71 GGGAAACGCTGAGCGCCCGAGGGCAGATGAGAGATGAGAGAGCCCTCGGGGT 130
QY 705 GGTGGCGGCTCGGAGGCTTCAGTCAAGCCGCTGCTCAACAGATCCTTGTGTCAA 764
DB 131 GGTGGCGGCTCGGAGGCTTCAGTCAAGCCGCTGCTCAACAGATCCTTGTGTCAA 190
QY 765 TTGGACCTCATGAAACAGACGACGACGCTGCAAGCCCAAGAAAGAGATCGAGAG 824
DB 191 TTGGACCTCATGAAACAGACGACGACGCTGCAAGCCCAAGAAAGAGATCGAGAG 250
QY 825 CTGAAGTCAGAGAGACGCTCTGCTCGGATTGGAAGTATGAAAGCGCGATGAC 884
DB 251 CTGAAGTCAGAGAGACGCTCTGCTCGGATTGGAAGTATGAAAGCGCGATGAC 310
QY 885 CTGTAAAGAGATTAAGAAAGAAAGGACACAGCTGTTTCAAGGCTATGAACTGAA 944
DB 311 CTGTAAAGAGATTAAGAAAGAAAGGACACAGCTGTTTCAAGGCTATGAACTGAA 370
QY 945 GAGAGAGAGAAACAGAGCTATCGAAGAAATTAACGTGAGTCCAGCGGAGCTTCC 1004
DB 371 GAGAGAGAGAAACAGAGCTATCGAAGAAATTAACGTGAGTCCAGCGGAGCTTCC 430
QY 1005 GAGACATCCAGACTGCTCTCCCAAGCCCTTCTCATGTGGCGGAGTGAAGAGCAT 1064
DB 431 GAGACATCCAGACTGCTCTCCCAAGCCCTTCTCATGTGGCGGAGTGAAGAGCAT 490
QY 1065 AAAAGAAATCCCATTTGGAAGTACGAAAGAAAGACTCTCTTAAAGAGTGGCTCT 1124
DB 491 AAAAGAAATCCCATTTGGAAGTACGAAAGAAAGACTCTCTTAAAGAGTGGCTCT 550
QY 1125 GAATTTCAAAAGTCAAAACAAAAACCTCTAAGCACTCTCTAATTAAGAGAACCTGT 1184
DB 551 GAATTTCAAAAGTCAAAACAAAAACCTCTAAGCACTCTCTAATTAAGAGAACCTGT 610
QY 1185 GGTTCCTTAATCTGAACTGTTTGTAAACGTGAATGAGAGCCAGAAACCCAGAAAG 1244
DB 611 GGTTCCTTAATCTGAACTGTTTGTAAACGTGAATGAGAGCCAGAAACCCAGAAAG 670
QY 1245 CCCCAGTCTTCAAGTGAACCCCAACCAAGACTCTCCACCCCAAAAGGAGCCAGACC 1304
DB 671 CCCCAGTCTTCAAGTGAACCCCAACCAAGACTCTCCACCCCAAAAGGAGCCAGACC 730
QY 1305 CATTCCAAAGAGAAAGCTTCTCAAGTGAATGAGATTTGCGGTACCTTTCAACCA 1364
DB 731 CATTCCAAAGAGAAAGCTTCTCAAGTGAATGAGATTTGCGGTACCTTTCAACCA 790
QY 1365 GAAATGATTTGTGCGTGGACACAGCTCCCCATCACCGTTACCATTAAGGGAATCC 1424
DB 791 GAAATGATTTGTGCGTGGACACAGCTCCCCATCACCGTTACCATTAAGGGAATCC 850
QY 1425 TCTTCAAGAGAGAGACTGTAGCAAGGTGTCTGATGCAATCAAGTGTGACAGAGA 1484
DB 851 TCTTCAAGAGAGAGACTGTAGCAAGGTGTCTGATGCAATCAAGTGTGACAGAGA 910
QY 1485 ACTTCAGTCTTGGCTGTCTTCTTGAAGGAGACCTCAAGTAAAGCTTAAAGGACCA 1544

DB 911 ACTTCAGTCTTGGCTGTCTTCTTCTTGAAGGACCACTCAAGTAAAGCTCTAAAGGACCA 970
QY 1545 AATCCTTCAACCTTTTGGAGAACTGATGACAGTGTGTTTGAAGCGGACATGCAAA 1604
DB 971 AACCTTTCAGACTTTTGGAGAACTGATGACAGTGTGTTTGAAGCGGACATGCAAA 1030
QY 1605 CTGAGACTGATGAGAGAGAGAAAGATGAGATTTTCAAGATTCAGGAAACAAAGA 1664
DB 1031 CTGAGACTGATGAGAGAGAGAAAGATGAGATTTTCAAGATTCAGGAAACAAAGA 1090
QY 1665 ATTTTACAGCACTGACGCTCAGAAATGTATTAAGAAAGAAATTCAGGAATCTGACCT 1724
DB 1091 ATTTTACAGCACTGACGCTCAGAAATGTATTAAGAAAGAAATTCAGGAATCTGACCT 1150
QY 1725 GAGGTACCTCATTTTTCCTGAGCCAGATGATGTTGAAGTTTGAATTAACCCCTTC 1784
DB 1151 GAGGTACCTCATTTTTCCTGAGCCAGATGATGTTGAAGTTTGAATTAACCCCTTC 1210
QY 1785 TTGCTGTGTGATGATTTGAGAGACATTAACCAAAATTAATTCACAGAAATTTGAGCTA 1844
DB 1211 TTGCTGTGTGATGATTTGAGAGACATTAACCAAAATTAATTCACAGAAATTTGAGCTA 1270
QY 1845 CCCTGTTGAGATGAGCGTACCGATGACATTTGAGATCCAGAAAGAAACACCTCAC 1904
DB 1271 CCCTGTTGAGATGAGCGTACCGATGACATTTGAGATCCAGAAAGAAACACCTCAC 1330
QY 1905 CGGACGTATGAGAAATAGCTGTGCGCAAGAACCTGTCTTCAGATGTTGATGATGC 1964
DB 1331 CGGACGTATGAGAAATAGCTGTGCGCAAGAACCTGTCTTCAGATGTTGATGATGC 1390
QY 1965 CATTCGCCGAGATGAGAGACCTGTATATGACCTTTGTCTCTCATATGTATCACT 2024
DB 1391 CATTCGCCGAGATGAGAGACCTGTATATGACCTTTGTCTCTCATATGTATCACT 1450
QY 2025 CGCTGATTAATACCTTTCATACCTTCTGACCTTGTTCATTAATCTGATTTCACAAA 2084
DB 1451 CGCTGATTAATACCTTTCATACCTTCTGACCTTGTTCATTAATCTGATTTCACAAA 1510
QY 2085 ACTCTTTCATTCGGCTAATGTGATGATTAAGAGGTGATTTGGATTTCTTTCCTTTT 2144
DB 1511 ACTCTTTCATTCGGCTAATGTGATGATTAAGAGGTGATTTGGATTTCTTTCCTTTT 1570
QY 2145 TGGAAATAGGCTCTTAAGCTAAGCTATAGATGAGCAATTCAGAAAGTTTCAAGGCTCT 2204
DB 1571 TGGAAATAGGCTCTTAAGCTAAGCTATAGATGAGCAATTCAGAAAGTTTCAAGGCTCT 1630
QY 2205 GTTTTATATCATTTGCTATGTTAAAGGGGTAAAGGGGCTCTTCAATTAAGACATGAG 2264
DB 1631 GTTTTATATCATTTGCTATGTTAAAGGGGTAAAGGGGCTCTTCAATTAAGACATGAG 1690
QY 2265 AGATGAAGAGCCCTTCTTCTTGAAGCTGTGCTGATGAGCACTTCTTCAACCTGTATC 2324
DB 1691 AGATGAAGAGCCCTTCTTCTTGAAGCTGTGCTGATGAGCACTTCTTCAACCTGTATC 1750
QY 2325 ACCCTCTTAATAGTGGTATAGTATTTTAAACCTTAATTAACAAACAACTCTACCA 2384
DB 1751 ACCCTCTTAATAGTGGTATAGTATTTTAAACCTTAATTAACAAACAACTCTACCA 1810
QY 2385 TGAAGCTTTGAGACCAAGAGAAATGACAGAGTGAAGCGATGAAGCAAGCATCTTACAG 2444
DB 1811 TGAAGCTTTGAGACCAAGAGAAATGACAGAGTGAAGCGATGAAGCAAGCATCTTACAG 1870
QY 2445 AGTGAAGAAAGACATCGAGAGTGTGATGATTAATCTGTCAAAAGATGTGTTCAATTGA 2504
DB 1871 AGTGAAGAAAGACATCGAGAGTGTGATGATTAATCTGTCAAAAGATGTGTTCAATTGA 1930
QY 2505 AACTATTCGTGATACAGTCAATGTGGAAGGAGATGTTGCTGTGTATATTTTTCAGTT 2564
DB 1931 AACTATTCGTGATACAGTCAATGTGGAAGGAGATGTTGCTGTGTATATTTTTCAGTT 1990
QY 2565 AATGATTAACATTTCTTATCTGCTCAAAACCAAAATCTTTTGAAGAAAGAAAGTGGGAT 2624
DB 1991 AATGATTAACATTTCTTATCTGCTCAAAACCAAAATCTTTTGAAGAAAGAAAGTGGGAT 2050

QY 2625 GGTAGCTTTGAGAACAGTTACAGCTGTAAACAAAAGCACTTAGTATTGGAGTGCATG 2684
| | | | |
Db 2051 GGTTAGTTTCAGAACAGTTACAGCTGTAAACAAAAGCACTTAGTATTGGAGTGCATG 2110
QY 2685 CCAAAACCTGTAAATATGCTCTGTATCAGATCACTTCTCAAGTATTCCTTCTTGGGCT 2744
| | | | |
Db 2111 CCAAAACCTGTAAATATGCTCTGTATCAGATCACTTCTCAAGTATTCCTTCTTGGGCT 2170
QY 2745 TCATCTCTTTAGAGAACTCTGTGTGTGGATAGAGACTTAGAGAGGGTGGGAGAG 2804
| | | | |
Db 2171 TCATCTCTTTAGAGAACTCTGTGTGTGGATAGAGACTTAGAGAGGGTGGGAGAG 2230
QY 2805 TGTGAAATAGTGTCTCTTGGCTGTGCAATGTCTACATCTTGAACCAACAGATGTA 2864
| | | | |
Db 2231 TGTGAAATAGTGTCTCTTGGCTGTGCAATGTCTACATCTTGAACCAACAGATGTA 2290
QY 2865 CCTAATAGCTTCTCCATTCATCTTGTAAATATTTGTATGTATCCATCTTGGCTCT 2924
| | | | |
Db 2291 CCTAATAGCTTCTCCATTCATCTTGTAAATATTTGTATGTATCCATCTTGGCTCT 2350
QY 2925 CTCCTCCCTCCGTTTGTAAATATGAGATAGACCTCCAGGCACTTGGCTGATG 2984
| | | | |
Db 2351 CTCCTCCCTCCGTTTGTAAATATGAGATAGACCTCCAGGCACTTGGCTGATG 2410
QY 2985 TAAGATCCCTATTAATCTGTAAGGAAATAGAGCAAGCCTGTGTCTCAATATA 3044
| | | | |
Db 2411 TAAGATCCCTATTAATCTGTAAGGAAATAGAGCAAGCCTGTGTCTCAATATA 2470
QY 3045 TAGAAATGCTCTTCTTGTAGCTTCAAGACTATGTGTGAAAACAAGTAGGGCTTAATC 3104
| | | | |
Db 2471 TAGAAATGCTCTTCTTGTAGCTTCAAGACTATGTGTGAAAACAAGTAGGGCTTAATC 2530
QY 3105 TCCTAATAGTAGGGCTTTATCCCTAAAGAAATATGTCCCAATTAATTAGCACTT 3164
| | | | |
Db 2531 TCCTAATAGTAGGGCTTTATCCCTAAAGAAATATGTCCCAATTAATTAGCACTT 2590
QY 3165 TAGAGAGAAAGCAAGATGTAGGGGTGTGGCTGGCCCATGATGAGACAGAAAGAG 3224
| | | | |
Db 2591 TAGAGAGAAAGCAAGATGTAGGGGTGTGGCTGGCCCATGATGAGACAGAAAGAG 2650
QY 3225 GAATGGGATACATTTGTGGAGAGAAAGAAAGTTCTCAAGGGGCTCCCACTGCTAAG 3284
| | | | |
Db 2651 GAATGGGATACATTTGTGGAGAGAAAGAAAGTTCTCAAGGGGCTCCCACTGCTAAG 2710
QY 3285 TTTTGTGTAGATGTGTATGTGTCTCTGTGATTTGACTTTTAAAGAAATATTTCTGGC 3344
| | | | |
Db 2711 TTTTGTGTAGATGTGTATGTGTCTCTGTGATTTGACTTTTAAAGAAATATTTCTGGC 2770
QY 3345 AGCAGATGTATCTTGTGATCTGTGCTGCTTATTTCTCTTTTGTGTGTG 3404
| | | | |
Db 2771 AGCAGATGTATCTTGTGATCTGTGCTGCTTATTTCTCTTTTGTGTGTG 2830
QY 3405 TGTGTGTGTGTGCTATGAGTTTCAATTGTAACTCCATCTGCTTGAAGAGTGGCTCT 3464
| | | | |
Db 2831 TGTGTGTGTGTGCTATGAGTTTCAATTGTAACTCCATCTGCTTGAAGAGTGGCTCT 2890
QY 3465 CTATTAAGGAACCTGCTGTAACTTCACTGACAGAGATGTAGAGAAATGTGACTTA 3524
| | | | |
Db 2891 CTATTAAGGAACCTGCTGTAACTTCACTGACAGAGATGTAGAGAAATGTGACTTA 2950
QY 3525 ATTCCAAGGGGCTGCTATCTACACCTTAAGAGAGAGATTTCTGAAGAAACCTGGGCCA 3584
| | | | |
Db 2951 ATTCCAAGGGGCTGCTATCTACACCTTAAGAGAGAGATTTCTGAAGAAACCTGGGCCA 3010
QY 3585 GATTTTCTTGTCTCCATCATTTTAATGTGAGAGGCTGTTCAGTTTCTTACTCTTACC 3644
| | | | |
Db 3011 GATTTTCTTGTCTCCATCATTTTAATGTGAGAGGCTGTTCAGTTTCTTACTCTTACC 3070
QY 3645 TATGTATATTTCTTGTATACGTGTCCAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3704
| | | | |
Db 3071 TATGTATATTTCTTGTATACGTGTCCAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3130

QY 3705 CTTTGTCTTTGATCCCTCAGATTCTTCTTGAATTTAGACATGTGTGGCTTCTAATTTT 3764
| | | | |
Db 3131 CTTTGTCTTTGATCCCTCAGATTCTTCTTGAATTTAGACATGTGTGGCTTCTAATTTT 3190
QY 3765 GGGTATGAGTTGCAAAATTAACCATTTGTGTGTGGCCCTACCCAGGGGCTCCAGTT 3824
| | | | |
Db 3191 GGGTATGAGTTGCAAAATTAACCATTTGTGTGTGGCCCTACCCAGGGGCTCCAGTT 3250
QY 3825 TCTGACTTAAGTAGACGTAAGAAATCCACGAGGTCTATCTGGCCAGATTTAAGTAGA 3884
| | | | |
Db 3251 TCTGACTTAAGTAGACGTAAGAAATCCACGAGGTCTATCTGGCCAGATTTAAGTAGA 3310
QY 3885 TTCTATTTCTCTGTGTCTCCCTCTCTCCCTAGACACTTATTTATTTATTTGCCCTTTCTA 3944
| | | | |
Db 3311 TTCTATTTCTCTGTGTCTCCCTCTCTCCCTAGACACTTATTTATTTATTTGCCCTTTCTA 3370
QY 3945 GGTAAATTTCTCTTGAATTTGACTTTGTGAAGAGAGGTGACAGTGAATAGCAAG 4004
| | | | |
Db 3371 GGTAAATTTCTCTTGAATTTGACTTTGTGAAGAGAGGTGACAGTGAATAGCAAG 3430
QY 4005 TTCCAAGTCAAAATTAACAGTGTGTAGAGTGTGGGGGAAATTAAGTCTTATTTTCCC 4064
| | | | |
Db 3431 TTCCAAGTCAAAATTAACAGTGTGTAGAGTGTGGGGGAAATTAAGTCTTATTTTCCC 3490
QY 4065 TACATGAGATACACA CTGTGAATTCATCTTCACTGAAGGCCCTGCACTTCTCTAAA 4124
| | | | |
Db 3491 TACATGAGATACACA CTGTGAATTCATCTTCACTGAAGGCCCTGCACTTCTCTAAA 3550
QY 4125 ACATAGTGTGTGTCTTTCTTTAACAAGTTTAACTAGTGTAAATTAATTAATTAAT 4184
| | | | |
Db 3551 ACATAGTGTGTGTCTTTCTTTAACAAGTTTAACTAGTGTAAATTAATTAATTAAT 3610
QY 4185 TGCTGTGTCTACTTCACTCACTTGTGTATGCACTTCAATTTGTGTGTGTGTA 4244
| | | | |
Db 3611 TGCTGTGTCTACTTCACTCACTTGTGTATGCACTTCAATTTGTGTGTGTGTA 3670
QY 4245 ATTCATTAATCTTTGATACATTTCTGATGTGTAATTTGGTGTCTTGTAAATCTTAT 4304
| | | | |
Db 3671 ATTCATTAATCTTTGATACATTTCTGATGTGTAATTTGGTGTCTTGTAAATCTTAT 3730
QY 4305 AAAGAGTCAATTTGTAATTAACCTATTTGGCTGTGTAATTAATTAATTAATTAAT 4359
| | | | |
Db 3731 AAAGAGTCAATTTGTAATTAACCTATTTGGCTGTGTAATTAATTAATTAATTAAT 3785
RESULT 2
AK014463
LOCUS 3161 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
library, clone:4121402D02 product:hypothetical protein, full insert
sequence.
ACCESSION AK014463
VERSION AK014463.1 GI:12852332
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99278253
PUBMED 10349636
REFERENCE
2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3

QY	775	TCGAACGACGACGACGACGCTGCAAGCCAAAGAAAAAGAGATCGAGAGCTGAAAGCAG	834
Db	1019	TCGAACGACGACGACGACGATTTGCAAGGCCAAAGGAAAGAGATGAGAGCTGAAATCCG	1078
QY	835	AGAGAGACAGCGCTCTCTTCCTCGATTGAACGTATGAGAAAGCGAGTCAGCTGTAAAGA	894
Db	1079	AGAGAGATAGCGCTCTCTTCCTCGATTGAACGTATGAGAAAGCGAGTCAGCTGTAAAGA	1138
QY	895	AGATTAACGAGAAAGAAAGGCAAGCTGTTCAAGGCTATGAATCGAAGAGAGAGG	954
Db	1139	GGGATTAACGAGAAAGAAAGGCAAGAGCTGCTCCAGGCTATGAATCGAAGAGAGAGG	1198
QY	955	AAAAGAGCTATCTGAGAAAAATTAACTGAGAGTCCAGCCGGACTTTCCGACATCCC	1014
Db	1199	AAGCAGAGTTGCTCGAGAAAATTAAATTGGAGCCCAAGCCGAGCTTTGCCAGACCTCCC	1258
QY	1015	AGACTGCGCCCTCCCAAGGCCCTTCATGTGGCGGAGTGGAGAAAGGACATTAAGAAAT	1074
Db	1259	AGGCTTGCTCTTCCAAAGCTTTCTCATGTGGCCGAGTGGAAAGGACCAAAAGGAAA	1318
QY	1075	CCCCATTTGGAGATGACAAAAGAAAGACTCTCTGTAAAAAAGCTGCTCTGAAATTTTCAA	1134
Db	1319	CCCCATTTGGAGAAATACAGAAAGAAAGAAATCTGTAAAAAGCTGCTCTGAAATTTTCAA	1378
QY	1135	AAGTCAAAACAAAACCTCTAAGACCTCTCTATTAAAGAGAAACCTGTGCTCTTAT	1194
Db	1379	AAGTCAAAACAAAACCTCTAAGACCTCTCTCAATTAAGAGAAACCTGTGCTCTAAT	1438
QY	1195	CTGAACCTGTTGTAACGTGAATTGAGAGACCAAGAAACCCCAAGAAAGCCCGGCTT	1254
Db	1439	CAGAACTGTTGTAACGTGAATTGAGAGACCAAGAAACCCCAAGAAAGCCCGGCTT	1498
QY	1255	CAGTGGACACCCCAACCAAGACTCTCACTCCCCAAAAGGAGCCAGCACCCATCCCAAG	1314
Db	1499	CAGTGGATCCCAACCAAGACTCTCGACTCCCCAAAAGGAGCCAGCACCCCAAG	1558
QY	1315	AGAAAGCCTTCTCAAGTAGAGATAGAAGATTTGCCGTACTTTCCACCAAGAAATGTATT	1374
Db	1559	AGAAAGCCTTCTCAAGTAGAGATAGAAGATTTGCCGTACTTTCCACCAAGAAATGTATT	1618
QY	1375	TGTGTCGTTGGCACCAAGCTTCCCCATACCGTTACCTTACGGGAATCCTCTCCAAAGA	1434
Db	1619	TGTGTCGTTGGCACCAAGCTTCCCCATACCGTTACCTTACGGGAATCCTCTCCAAAGA	1678
QY	1435	AGGAGAGACCTGTAGCAAGGTGTGATAGCCATCAAGTTGACAGAGAAACCTCAGCT	1494
Db	1679	AGGAGAGACCTGTAGCAAGGTGTGATAGCCATCAAGTTGACAGAGAAACCTCAGCT	1738
QY	1495	TGCGCTGTTCTCTTCTTGAGGGACCACTCAGTAGAGCCTCTAAGGGACCCCAATCCTCAG	1554
Db	1739	TGCGCTGTTCTCTTCTTGAGGGACCACTCAGTAGAGCCTCTAAGGGACCCCAATCCTCAG	1798
QY	1555	ACCTTTTGGAGAACCTGGATGACAGTGTGTTTGGAGCGGCATGCAAACTGAGCTGG	1614
Db	1799	ACATTTTGGAGAACCTGGATGACAGTGTATTTTCAAGAGCATGCAAAACTGAGCTAG	1858
QY	1615	ATGAGAGAGAGAGAAAGATGGGATATTCAAGAGATATAGGGAACAAAGAAATTTTACAGC	1674
Db	1859	ACGAGAGAGAGAGAAAGATGGGATATTCAAGAGATATAGGGAACAAAGAAATTTTACAGC	1918
QY	1675	GACGTCAGCTCAGATGTATATAAAAGAAAGAAATTCAGAGATCGAGCTCAGGTTACT	1734
Db	1919	GACGTCAGCTCAGATGTATATAAAAGAGAAATTCAGAGATTCGAGCTCAGGTTACT	1978
QY	1735	CATTTTTCCTGAGCCAGATGATGTGAAGATTGATGATTACCCCTTCTTGCCCTGTG	1794
Db	1979	CATTTTTCCTGAGCCAGATGATGTGAAGATTGCTGATTACCCCTTCTTGCCCTGTG	2038
QY	1795	TAGCATTTTGGACGACCAATTACCAAAATTAATCTCCACAAATTTTGAAGTACCCGTGG	1854
Db	2039	TAGCATTTTGGACGCGCATTTGCCAAATTAATGCTCCACAACTTTGAACTTACCCTGTGG	2098

QY	1855	ATGAGCGTACCCGATGCAAGATTGGAGATCCAGAAAGAAAGCAAAACCTTCAACCGGACGGTGA	1914
Db	2099	ATGAGCCAAACCGTTGACGGTTGGAGATCCAGAAAGAAAGCAAAACCTTCAACCGGACCGTGA	2158
QY	1915	GGAATAGCTGTGTGCGCAAAACCCTGTCTTCAGATAGTTGAGCAATGCCATTCC - CGA	1973
Db	2159	GGAAGTAGCTGTGTGCGC - GGTAGCTGTCTTCAGATAGTTGAGCAATGCCATTCCGAGA	2217
QY	1974	GAGTGGCAGAGACCTGTATATGTGACCTTTGTCTCTCACAATATGTTATCACTCGTGATTA	2033
Db	2218	GAATAGCAGAGACCTGTATATGTGACCTGTGCGCTTCAATCTGTATATCACTCGGTAA	2277
QY	2034	TACCCCTTCATACCTCTTGA-----CTTTGTTTTCATACCTGTATTCACAA-----	2082
Db	2278	TGCCTTTCATACCTCTTATATATAGTTTGTGTTCTTTTACCTGTATTCACAAACAA	2337
QY	2083	-----AAACTCTTTCATTCCGCTAATTTGATTTGAGAGGGTGAATGGGATTTCTTT	2135
Db	2338	ACAGAAAGAAAGCTTTCGCTTGCTGTATTTGTGAATTA---GGGTTACTGGAGCGCTTT	2394
QY	2136	TCCCTTTTTTTGGAAATGCGCTCTCAAGCTTAAAGCTATAGAGATGGCAGATTTCAGAAATT	2195
Db	2395	TCCCAATTTTGGGAGAGATAG--CTTAGACTGTGTCTATAGS--GGGAGACTTGGGGGTCC	2450
QY	2196	CAGGGGTCTGTTCTATATACATTTGGCTTATGTATTAAGGGGTAAAGGGCTCTTTCATTAG	2255
Db	2451	CAGGGGCACTGTGTATCTACCTGCTGCTTACCGGGGTAA---GGGCTCTTTCATTAG	2506
QY	2256	ACATGTGGAAGATGAAAGCAGCCCTTCTTATAGAGCTGCTGCTGACATGGCACTCTTCTCA	2315
Db	2507	-CGTGTCTAAGATGGAACAACTTCTGTATGAGAGCTGTGCTGCTGACCCCGTCCCT	2565
QY	2316	CCCTGTGACACCCCTCTTATAGTGGGTATAGTGAATTTTAAACCTTAAATTAACAACA	2375
Db	2566	CCCTGTGCGCTTCTGAATATGTGTGGTGTAGTAGT-----	2604
QY	2376	ACTTCACCAATGAGCTTTAGACCAAGAAAGGAATGACAAATGAAAGCATGATGACCA	2435
Db	2605	---TCACCGTGAATTTTATGGGCCCGGAGAAATTA---GCATGAAGTATGAGCAACACA	2656
QY	2436	TCTTCACAGATGAAGAAAGCATGCGAGAGTGGTATATATACGTCTGAAAGATAGTTG	2495
Db	2657	TTCTCTCAGAGGGGAAACTGCGAG-----TTTATCTGAAAGTTTG	2698
QY	2496	TTCAATTTGAAACTATTTCTGTGATTCACATCATGTGGAAAGGAGATTTGGCTGTGATATT	2555
Db	2699	TTCTCTTAAACTGCTCTGTGACAGTGGCATGTGGAGAGACGTGACGTG - GTTCACT	2757
QY	2556	TTTTCACTTAATGATTAAC--AATTTCTTACTGTCTCAAAAACCAAAATCTTTGGAAA	2612
Db	2758	TTCTTAGTGTATGGGTAACTGAAGTTCTTCCACCCCAACAAGAAAGATGTTTATGAAA	2817
QY	2613	GAAAGTGGGATGTTAGTTTCAGAACAGTACAGCTGTAAA - CAAGAACAATTAGTAT	2671
Db	2818	AGAAATTTGGGATGTAAAGTTTCCAACTAGTTAAACAGTCTAAACCGATPAGAGCTTCATT	2877
QY	2672	TTGGAGTGGCATGTCAAAAACCTGTATTAATAGTCCCTGTATACACATCACTTCTCAAGATT	2731
Db	2878	TGGGATGCGCATGCCAAGATCTGTAGAGATGTCTCTG--TCAATCACTTTTCAAGATT	2935
QY	2732	CCTTCAATTTGGCTTCACTCTTTTATAGCAAACTTGTGTGGTGGGATAGAGACTTGGGAG	2791
Db	2936	CTTTATGTGGGCTTGATCTCTATTATGCAAACTTTTGGTGGGATGTGAGATTTGGGGTG	2995
QY	2792	GG--TAGGGGAGAGTGTGAAAATAGTGTCTTCTTTGGCTGGCAATGTCTACATTTTG	2849
Db	2996	GGGCGGGGGATGGGTATGAAAGATATGCTTCTGTGCTGGCGAGATGTCTACAGCTTG	3055
QY	2850	AAACAAACAGATGTACCTAATGAGCTTCTCCATTCACATTTGTGAAAATTA - ATTTGTATGT	2908
Db	3056	AAACAAACAGTGTACTTAATGAACCTT--TATTCGTTATGTGAATAATAGTTTATACAC	3112
QY	2909	GTACCACTTGTGTCCTCTCCCTCCGCTTTTGT	2942

Db	Accession	Definition	Accession	Definition
RESULT 3	AK053719	2839 bp mRNA linear HTC 03-APR-2004		
LOCUS	AK053719			
DEFINITION	Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched library, clone:EI3032509 product:hypothetical protein, full insert sequence.			
ACCESSION	AK053719			
VERSION	AK053719.1	GI:26343664		
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	1 Carninci, P. and Hayashizaki, Y.			
AUTHORS	High-efficiency full-length cDNA cloning			
TITLE	Meth. Enzymol. 303, 19-44 (1999)			
JOURNAL	99279253			
MEDLINE	10349636			
PUBMED	2			
REFERENCE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komu, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
AUTHORS	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
TITLE	Genome Res. 10 (10), 1617-1630 (2000)			
JOURNAL	20499374			
MEDLINE	11042159			
PUBMED	3			
REFERENCE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komu, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Obara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.			
AUTHORS	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer			
TITLE	Genome Res. 10 (11), 1757-1771 (2000)			
JOURNAL	20530913			
MEDLINE	11076861			
PUBMED	4			
REFERENCE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.			
AUTHORS	Functional annotation of a full-length mouse cDNA collection			
TITLE	Nature 409, 685-690 (2001)			
JOURNAL	5			
MEDLINE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.			
PUBMED	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs			
REFERENCE	Nature 420, 563-573 (2002)			
AUTHORS	6 (bases 1 to 2839)			
TITLE	Adachi, J., Aizawa, K., Akiyama, T., Hara, A., Hashimoto, W., Fukuda, S., Furuno, M., Hanagaki, T., Harada, A., Hayashizaki, K., Hayatsu, N., Hiramoto, K., Hirose, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komu, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akihira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.			
AUTHORS	Direct Submission			
TITLE	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,			

Kanagawa 230-0045, Japan (E-mail: genome-res@gsic.riken.jp, URL: http://genome.gsic.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.
 URL: http://genome.gsic.riken.jp/
 URL: http://fantom.gsic.riken.jp/
 Location/Qualifiers

```

1. 2839
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="C57BL/6J"
    /db_xref="FANTOM_DB:EI30302B09"
    /db_xref="taxon:10090"
    /clone="EI30302B09"
    /tissue_type="eyeball"
    /clone_1fb="RIKEN full-length enriched mouse cDNA library"
    /dev_stage="0 day neonate"
33. 1829
    /note="unnamed protein product; hypothetical protein
(evidence: Procrest)
putative"
    /codon_start=1
    /protein_id="BAC35489.1"
    /db_xref="GI:26343665"
    /translation="MTMRASVFRPAAAPAGGNPDRLDYERAAALGGPEDESGAAEAH
FLRPHRLTKEPPEPLASOGSPSPRSGGSGKGRLLPAGAAPEQGESNMGSPV
LPCEPRTKQAGIGGEVVAAGAGSPSPKQAVLPITGSI VAAAAEPTPWADKGG
AAPAAVATSDPAGPPLPLPGPPPLATAGATLAAEGMTKSIRKSLGGGSGAS
SOAAKLIQIILLDLIEQQQDLQAAEKIIEIKSERDYLAIIRIEMERMOVLKSD
NPKRKHLLQGVGEEREEBALESEKILGSEOPELCESOALPSKPFSGSGSGKHKR
TPGNTERTKTPVKKLAPEERKVTGPKHSPIKEEPGSISEPTKCELASOETPEKR
RSSVDTPRLSTPOKGPSTIPKEXAFSSSEMDLPYLSTTEMYLCRMHQPPSPPLRE
SSPEKERTVALPSWRHDSVEPLDPNPSDILLENLDVSFKRHAKEILDEKRRRWID
QRIREKTIILRIQLRMVKKGIQIESPEEVSFPPEPDVESLITPLPVAAFGRPPLP
KLAPQNFELPWLDEAKPLQVGPDEAHTSPDL"
```

ORIGIN

Query Match	41.1%	Score 1798.6	DB 3	Length 2839
Best Local Similarity	82.0%	Pred. No. 0		
Matches 2403	Conservative 0	Mismatches 379	Indels 147	Gaps 23
47	GCTCGGTGTCGCGCGGTGCTCGCGACACATGATGATCGCGGTTCAGAGC	106		
Db	2	GTCGTGGTCCCCCGCGCTGCTCGGACACATATACATGAGTCGAGTTCAGAGC	61	
47	GACCGCGAGCCCTTCGCGCGGCGATCCTGAGCAGAGCTGAGCTGAGCGGCTGCGGC	166		
Db	62	GCGCGCGGCCCCCTTCGCGCGGCGAACCCCGAGCAGCACTGAGCTACGAGCGGCTGCGGC	121	
47	167	GCTGGGCGGCGCGGAGCAGAGCTTGGGCGGCGCGGAGCCCATCTTCTCCCGCGGACCG	226	
Db	122	GCTGGGCGGCGCGGAGCAGAGCTTGGGCGGCGCGGAGCCCATCTTCTCCCGCGGACCG	181	
47	227	TAACTCAAGAGCGCGCGGCGGCGGCTCTCCCAAGGCGGAGCGCCCGGCGCTTC	286	
Db	182	TAACTCAAGAGCGCGCGGCGGCGGCTCTTCCTCAAGGCGGAGCGCCCGGCGCTTC	241	
47	287	CCCGGCGGCGCT---GCGGCGGCGAGGCGCGGCGGCTTGTAATCCCGGCGGCGGCGGCGC	343	
Db	242	TTCAGCGGCGCTGCGGCGGCGGCGAGGCGCGGCGCTTGTAATCCCGAGCGGCGGCGC	301	
47	344	CGGCGACAGAGAGAGCTTGGGCGGCTTGGTGGTCCCTTGCCTTCGCGCCCGCGGCAC	403	
Db	302	CGGCGACAGAGAGAGAGCTTGGGCGGCTTGGTGGTCCCTTGCCTTCGCGCCCGCGGCAC	361	
47	404	CAGAGCAAGCGCGGCACTTGGGCGGCGAGCTGCCGCGAGCGGAGCGGCTGCAAGCGCGCGC	463	
Db	362	TAAACAGCGCGGCACTGCGCGGCGGCGAGCTGCGAGCGGCGGCTGCAAGCGCGCGCGC	421	

QY 464 CAAGTATCAGCGGTGTCCTCCATTGACAGCGGCTCTCTC---GTGGCGGCGCCAAAGA 520
Db 422 CAAGTATCAGCGGTGTCCTCCATTGACAGCGGCTCTATGATGTGGCGGCGCAAGA 481
QY 521 GCTTAGCGCTGGGCTGGGGAACAAGGTGGGGCGGCTCCCCGCTGCCACCGCTCGGA 580
Db 482 GCTTAGCGCTGGGCTGGGGAACAAGGTGGGGCGGCTCCCCGCTGCCACCGCTCGGA 541
QY 581 CCGGCGGGAAGCCCAACCACTACCTTGCCTGGGCGGCAAGCCCTGGCGGCAAGCGCAC 640
Db 542 CCGGCGGGAAGCCCAACCACTACCTTGCCTGGGCGGCAAGCCCTGGCGGCAAGCGCAC 601
QY 641 CCGCGGGAAGCCCTGGCGGCAAGCGGCAAGGTGGAAGATGAGAGAGAGCCCTCTCG 700
Db 602 TGTGGGGAAGCCCTGGCGGCAAGGTGGAAGATGAGAGAGAGAGCCCTCTCG 661
QY 701 GGGTGTGGCGGCTGGGAGCTTCAGTCAAGCGGCTGCTCAACAGATCTTGTCT 760
Db 662 GGGTGTGGCGGCTGGGAGCTTCAGTCAAGCGGCTGCTCAACAGATCTTGTCT 721
QY 761 GCAATTGGAAGCTGATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 820
Db 722 GCAATTGGAAGCTGATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 781
QY 821 GAGCTGGAAGTCAAGAT 880
Db 782 GAGCTGGAAGTCAAGAT 841
QY 881 GAGCTGGAAGAT 940
Db 842 GAGCTGGAAGAT 901
QY 941 TGAAGAT 1000
Db 902 TGAAGAT 961
QY 1001 TTCCGAT 1060
Db 962 TTCCGAT 1021
QY 1061 ACATTAAGAT 1120
Db 1022 ACATTAAGAT 1081
QY 1121 TCCGTGAATTTTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCA 1180
Db 1082 TCCGTGAATTTTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCA 1141
QY 1181 CTGTGATTCCTTAATCTGAACTGTTTGTAAAGTGAATGAGAGAGAGAGAGAGAT 1240
Db 1142 CTGTGATTCCTTAATCTGAACTGTTTGTAAAGTGAATGAGAGAGAGAGAGAGAT 1201
QY 1241 AAAGCCCGGCTTTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1300
Db 1202 AAAGCCCGGCTTTCAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1261
QY 1301 CAGCCATCCCAAGAT 1360
Db 1262 CAGCCATCCCAAGAT 1321
QY 1361 CAGAGAAAGTGAATTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1420
Db 1322 CAGAGAAAGTGAATTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1381
QY 1421 ATCTCTCAAAAGAT 1480
Db 1382 ATCTCTCAAAAGAT 1443
QY 1481 AGAAACTTCAAGTCTTGGCTGTTCTTCTTGAAGAGAGAGAGAGAGAGAGAT 1540
Db 1414 AGAAACTTCAAGTCTTGGCTGTTCTTCTTGAAGAGAGAGAGAGAGAGAGAT 1453

QY 1541 CCCAAATCTTACAGCTTTTGGAGAACTGGATGACAGTGTGTTTCCAGCGGATGC 1600
Db 1454 CCCAAATCTTACAGCTTTTGGAGAACTGGATGACAGTGTGTTTCCAGCGGATGC 1513
QY 1601 AAAAATGAT 1660
Db 1514 AAAAATGAT 1573
QY 1661 AAAAATGAT 1720
Db 1574 AAAAATGAT 1633
QY 1721 GCTGAGAGTGAATCTATTTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAT 1780
Db 1634 GCTGAGAGTGAATCTATTTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAT 1693
QY 1781 CTCTGAGAGTGAATCTATTTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAT 1840
Db 1694 CTCTGAGAGTGAATCTATTTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAT 1753
QY 1841 GCTGAGAGTGAATCTATTTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAT 1899
Db 1754 GCTGAGAGTGAATCTATTTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAT 1813
QY 1900 CTCACCGAGAGTGAATCTATTTTCCCTGAGAGAGAGAGAGAGAGAGAGAT 1959
Db 1814 CTCACCGAGAGTGAATCTATTTTCCCTGAGAGAGAGAGAGAGAGAGAGAT 1872
QY 1960 CATGCAATTC-CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2018
Db 1873 CATGCAATTC-CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1932
QY 2019 ATCACTGAGATTAATCCCTTATATCTTCTGAGAGAGAGAGAGAGAGAT 2072
Db 1933 ATCACTGAGATTAATCCCTTATATCTTCTGAGAGAGAGAGAGAGAGAT 1992
QY 2073 GATTTCAGAA-----AAACTCTTCAATGAGAGAGAGAGAGAGAT 2120
Db 1993 GATTTCAGAA-----AAACTCTTCAATGAGAGAGAGAGAGAGAT 2049
QY 2121 GATTGAGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2180
Db 2050 GATTGAGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2105
QY 2181 CAGATTCAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2240
Db 2106 CAGATTCAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2162
QY 2241 GGGCTCTCTTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2300
Db 2163 GGGCTCTCTTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2220
QY 2301 ATGCACTCTTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2360
Db 2221 CTGAT 2274
QY 2361 AAAAATGAT 2420
Db 2275 AAAAATGAT 2311
QY 2421 CGATGAT 2480
Db 2312 CGATGAT 2353
QY 2481 CTGAT 2540
Db 2354 CTGAT 2413
QY 2541 TTGGCTGATGATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2597
Db 2414 GAT 2472
QY 2598 AAAATCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2656

Db	2473	GAAATGTTTGGAAAAGAGATTGGGAGCTGTAGTTTCCAACTAGCTTAACAGTCTAAACC	25322
QY	2657	AAAAACATTAGTATTTGGGATGGCATGCCAAAACCTGTATAAATGTCCTGTATACAT	2716
Db	2533	GATAGAGCTTCATTTGGGGATGGCATGCCAGAGCTGTAGAGATGTC--TCTGCAAGT	2590
QY	2717	CACCTTCACGATATTCCTTCATTTGGGCTTCATCCCTTTAGCAAGAACTCTTGATGGGA	2776
Db	2591	CACCTTTCAGAGATTCCTTAAGTGGGGCTGTAGTCTATTAGCCAACTTTTGATGGTAGG	2650
QY	2777	TAGAGACTTAGGAGAGG--TAGGGGGAGAGTGTGGAAAATAGTGTCTCTTTGGCTGGCA	2834
Db	2651	TAGAGATTTGGGGTGGGGCCGGGGGTAGGATATGGAAGAAATGCTTCTCTGCTGGCA	2710
QY	2835	AATGCTACATCTTTGAAAACAACAGATGTACCTAATGAGCTTTCATTCACATTCATTA	2894
Db	2711	GATGCTACAGCTTGAACAACAACAGTTGTACTATGAACATT--TATTCGTTAGTGA	2767
QY	2895	AATA-ATTGTAGTGTACCATCTTGGTCTCTCCCTCCCGCTTTGGT	2942
Db	2768	AATGATTTATACAGAACCATCTTGTTTCCATATCTCTGCCTTGTT	2816

RESULT 4	
LOCUS	AK015496
DEFINITION	AK015496 1304 bp mRNA linear HTC 03-APR-2004
ACCESSION	AK015496
VERSION	AK015496.1
KEYWORDS	mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930463F05 product:hypothetical protein, full insert sequence.
SOURCE	AK015496 1 GI:12853861
ORGANISM	HTC; CAP trapper.
	Mus musculus (house mouse)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 Carninci, P. and Hayashizaki, Y.
AUTHORS	High-efficiency full-length cDNA cloning
TITLE	Meth. Enzymol. 303, 19-44 (1999)
JOURNAL	99279253
MEDLINE	10349636
PUBMED	
REFERENCE	2
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagoaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, U., Nishi, K., Kitanuma, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishimune, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format
JOURNAL	sequencing pipeline with 384 multicapillary sequencer
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	20530913
REFERENCE	11076861
AUTHORS	4
	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
MEDLINE	5
PUBMED	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
REFERENCE	Analysis of the mouse transcriptome based on functional annotation
TITLE	

JOURNAL OF 60, 770 full-length cDNAs
REFERENCE Nature 420, 563-573 (2002)
AUTHORS (bases 1 to 1304)
Aachai,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukuishi,Y.,
Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
Hirooka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
Kasukawa,T., Kato,H., Kawai,J., Koijima,Y., Komoto,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishii,K.,
Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D.,
Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT Please visit our web site (<http://genome.gsc.riken.jp/>) for further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAAGACAGATCCAGACCTCTTTTTCCTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAAGATTCGAGTTAAATTGAATTAATGCCCCCCCCC 3']. cDNA was cleaved
with BamHI and XhoI. cDNA of size comigrated longer than 7 kb was
selected before cloning. Vector: a modified pBluescript KS(+) after
bulk excision from Lambda PLC 1. Cloning sites, 5' end: SalI, 3'
end: BamHI. Host: DH10B.

CDS
location/Qualifiers
1..1304
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:4930463F05"
 /db_xref="taxon:10090"
 /cclone="4930463F05"
 /sex="male"
 /tissue_type="testis"
 /clone_id="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
17..129
 /note="unnamed protein product; hypothetical protein
 (evidence: ProCrest)"
putative"
 /codon_start=1
 /protein_id="BAB29868.1"
 /db_xref="GI:12853862"
translation="MLRFRARHGQALIPSLAQTTTNRNKLARIERMRRQLVLR
RDNKEHRHLQGYEPEREAEALISEKIQLERPELCETSDQALSPKFGSGSGKH
RKTPFGNTERKTPIVKLAPESFKVTXTPPKPISKEBPISITVCWRELRSDEPB
KSSPVDTPLRLSTPGKSTPHKEKAFSEMEDLPYLSTTEMICMHPDPPLPLD
REPSSKEETVAIPSWMDHSVEPLRDNPSPDIENLDSDVSFKHALTEDEKKRWK
DIQRIRQRILIQILAMTKKGIGOSEPPTVSFPEDVESILTRPLPYAVAFGRPR
LPKLAPONFLPWIDERSRCRLLEIQTAKTHRTCRK"

ORIGIN
Query Match 22.4% Score 980.8; DB 3; Length 1304;
Best Local Similarity 88.8%; Pred. No. 1,6e-205;
Matches 1129; Conservative 0; Mismatches 87; Indels 56; Gaps 4;

820 AGGAGTGAAAGTCAGAGAGACAAGCGCTCTGTTCGGATTGAACGTATGGAAGGCAGA 879


```
Db      121 TTTGATACCCATCTGCTTAGAGAGAGTGGGCTCTATATAAGGAACTGCTGTAACCTTC 180
Qy      3491 ATTGCAGCAAGATGTAGAGAGAAATAGACCTTAATTCCACTAGGAGCTCTCATCTCAACA 3550
Db      181 ATTGCAGCAAGATGTAGAGAGAAATAGACCTTAATTCCACTAGGAGCTCTCATCTCAACA 240
Qy      3551 CCTTAAGAGAGATTTCTAGAAAAAATCGGGCCAGATTTCTTGTGTTCCATCATTTTA 3610
Db      241 CCTTAAGAGAGATTTCTAGAAAAAATCGGGCCAGATTTCTTGTGTTCCATCATTTTA 300
Qy      3611 ATGTGGCAGGCTGTTCAGTTCCTTACTCTTACCTATGTGATATTTCTTCTGTACGTGTC 3670
Db      301 ATGTGGCAGGCTGTTCAGTTCCTTACTCTTACCTATGTGATATTTCTTCTGTACGTGTC 360
Qy      3671 CAAAAAGAAAAAGACCCATCATGCTCTCTTGAAGCTTTGTTCTTTGATCCCTCAGTTTCT 3730
Db      361 CAAAAAGAAAAAGACCCATCATGCTCTCTTGAAGCTTTGTTCTTTGATCCCTCAGTTTCT 420
Qy      3731 TCTTGATTTGAGATGTGCGGGTTCCTAATTTTGGGTATGAGTTAGCAATTTTAACAT 3790
Db      421 TCTTGATTTGAGATGTGCGGGTTCCTAATTTTGGGTATGAGTTAGCAATTTTAACAT 480
Qy      3791 TGTGTTTGTGCTTACCCAGGAGACTCCCAAGTTTCTGACTTGAAGTAGACTGAGAAGAA 3850
Db      481 TGTGTTTGTGCTTACCCAGGAGACTCCCAAGTTTCTGACTTGAAGTAGACTGAGAAGAA 540
Qy      3851 TCCACGAGGTGCTATCTGCGCAGATTTAAGTAGATTTCTAATTTCTTGGTTCTCCCTCTCC 3910
Db      541 TCCACGAGGTGCTATCTGCGCAGATTTAAGTAGATTTCTAATTTCTTGGTTCTCCCTCTCC 600
Qy      3911 CTGAGAGACCTCTTATTTTATGTCCTCTCTTCAAGTAAATTTCTCTTGAATTTGACTTT 3970
Db      601 CTGAGAGACCTCTTATTTTATGTCCTCTCTTCAAGTAAATTTCTCTTGAATTTGACTTT 660
Qy      3971 GTTGAGAGAGAGGTGAGACAGTAGATTAGCAAAAGTTCCAAAGTCAAAATTTACAGTGTT 4030
Db      661 GTTGAGAGAGAGGTGAGACAGTAGATTAGCAAAAGTTCCAAAGTCAAAATTTACAGTGTT 720
Qy      4031 AGAGTGTGGGGGAAAAATTAAGTCTTAATTTTCCCTACATGAGATACACAATCTGTAATTC 4090
Db      721 AGAGTGTGGGGGAAAAATTAAGTCTTAATTTTCCCTACATGAGATACACAATCTGTAATTC 780
Qy      4091 AATCTTCACTGAAGGCCCTGCGAGTCTCTTAATAACATAGTGTGTTGTTTCTTTAAACA 4150
Db      781 AATCTTCACTGAAGGCCCTGCGAGTCTCTCT-ANACATAGNTGTTGTTTCTTTAAACA 839
Qy      4151 AAGTTTAACTAGTGTAAATTAATAAAAAAATGCTGTCTGTCTACTGACGCTTGT 4210
Db      840 AAGTTTAACTAGTGTAAATTAATAAAAAAATGCTGTCTGTCTACTGACGCTTGT 899
Qy      4211 TTTATGCCCATTTTCATATTTGTTGTC--TGTGTGTATTCATCA-TCACCTTTGATACATTT 4267
Db      900 TTTATGCCCATTTTCATATTTGNTGNCCTGGGGTTGAATTCATTAACCTTTGATACATTT 959
Qy      4268 CTGATGTG--TAAATTTGTTGTTCTTTGTAATATCTTATTAAGAAGTTCAA-TTGAAT 4323
Db      960 CTGATGTG--TAAATTTGTTGTTCTTTGTAATATCTTATTAAGAAGTTCAA-TTGAAT 1019
Qy      4324 AACTATTTGTTGCTTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAA 4372
Db      1020 AACTATTTGTTGCTTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAA 1068
```

```
RESULT 6
BX329236/c 896 bp mRNA linear EST 08-APR-2004
LOCUS BX329236 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION CDNA clone CS0DK002YN11 3-PRIME, mRNA sequence.
ACCESSION BX329236
VERSION BX329236.2 GI:46280783
KEYWORDS EST.
SOURCE Homo sapiens (human)
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 896)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 2, 2003 this sequence version replaced gi:30344795.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with NotI and cloned
into the NotI and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
8989.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BA1036ZB04_CS03386_1&c=8989.f

FEATURES
source Location/Qualifiers
1..896
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK002YN11"
/cell_line="HELA"
/cell_line="HELA CELLS COT 25-NORMALIZED"
/clone_1ib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with NotI and cloned into the NotI and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 18.4%; Score 806.6; DB 5; Length 896;
Best Local Similarity 96.5%; Pred. No. 4,9e-167;
Matches 863; Conservative 0; Mismatches 26; Indels 5; Gaps 4;
Qy 303 GGCAGAGGCGCGGGCTTTGTAATCCCGCGCGGGGCGCCCCGGGCGAGCAAGAGAGAGC 362
Db 896 GGCAGAGGCGCGGGCTT-CTTACTCCGNNCGGGGCGCGCCCCGGGCGAGCAAGAGAGAGC 839
Qy 363 TGGGGCGTTTCGTCGCTTGTCCGTCGCGCCCGCGGCACCAAGCAAGCGGCATTGGG 422
Db 838 TGGGNCGG-TCGGTGCCCCCTGCTGTGCG-CCCGGGGACCAAGCAAGCGGCATG 782
Qy 423 GGGGAGCTCCCGCAGCGCGAGACCGGCTGACACCCCGGCGCAAGTATGAGCGGTG 482
Db 781 GGGGAGCTCCCGCAGCGCGAGACCGGCTGACACCCCGGCGCAAGTATGAGCGGTG 722
Qy 483 CCAATTCAAGCGGCTCTCTGTCGCGCGGCGCAAGAGCTTACGCTTGAGCTGGGAGAC 542
Db 721 CCAATTCAAGCGGCTCTCTGTCGCGCGGCGCAAGAGCTTACGCTTGAGCTGGGAGAC 662
Qy 543 AAGGTTGGGCGGCGCTCCCGCGCTGCGACCGGCTGCGACCGGCGGAGCCGCCACATTA 602
Db 661 AAGGTTGGGCGGCGCTCCCGCGCTGCGACCGGCTGCGACCGGCGGAGCCGCCACATTA 602
Qy 603 CCTGTGCGCGGCGCGCACCTCTCGCGGCCACCGGCACCGCGGAGCCTTGCGGCGACG 662
Db 601 CCTGTGCGCGGCGCGCACCTCTCGCGGCCACCGGCACCGCGGAGCCTTGCGGCGACG 542
Qy 663 GAGGCGAGTGAAGAGTATGAGAAAGAGCGCTTCTGGGGGTGGTGGGCGGCTCGGGAGCC 722
Db 541 GAGGCGAGTGAAGAGTATGAGAAAGAGCGCTTCTGGGGGTGGTGGGCGGCTCGGGAGCC 482
Qy 723 TCCAGTCAGGCGCGCTGCTCAACAGATCTTCTGCTCAATTGGAATCTCATCGAAGAC 782
Db 481 TCCAGTCAGGCGCGCTGCTCAACAGATCTTCTGCTCAATTGGAATCTCATCGAAGAC 422
Qy 783 CAGCAGCAGCAGCTGCGAGGCCAAGAAAAAGAGATTCAGAGAGCTGAATGTCAGAGAGAC 842
```

Db 421 CAGAGAGAGAGCTGAGGCGCAAGAAAGATCAGAGCTGAAGTCAGAGAGAC 362
Qy 843 ACCTCTCTTGTCTGATTTGAACGATGAAAGCGGATGCGAGTGTAAAGAGATAC 902
Db 361 ACCTCTCTTGTCTGATTTGAACGATGAAAGCGGATGCGAGTGTAAAGAGATAC 302
Qy 903 GAGAAAGAAAGCAAGAGCTGTTTCAAGGCTATGAAACTGAAAGAGAGAAACAGAG 962
Db 301 GAGAAAGAAAGCAAGAGCTGTTTCAAGGCTATGAAACTGAAAGAGAGAAACAGAG 242
Qy 963 CTATCTGAGAAATTAACCTGAGATGCCAGCGGAGCTTTCCGAGACATCCCACTCTG 1022
Db 241 CTATCTGAGAAATTAACCTGAGATGCCAGCGGAGCTTTCCGAGACATCCCACTCTG 182
Qy 1023 CTTCCCAAGCCCTCTCATGTCGGCGAGTGAAGGACATAAAGAAATCCCAATTT 1082
Db 181 CTTCCCAAGCCCTCTCATGTCGGCGAGTGAAGGACATAAAGAAATCCCAATTT 122
Qy 1083 GGAAGTACAGAAAGAAAGACTCTCTGTAAAGAGCTGCTCTGAATTTTCAAAAGTCAA 1142
Db 121 GGAAGTACAGAAAGAAAGACTCTCTGTAAAGAGCTGCTCTGAATTTTCAAAAGTCAA 62
Qy 1143 ACAAAACTCTTAAGCACTCTCTATTAAAGAGAACCTGTGTTCTTATCT 1196
Db 61 ACAAAACTCTTAAGCACTCTCTATTAAAGAGAACCTGTGTTCTTATCT 8

RESULT 7
BU857520 927 bp mRNA linear EST 16-OCT-2002
LOCUS AGENCOURT 10473579 NIH_MGC_107 Homo sapiens cDNA clone
DEFINITION IMAGE:6647374 5', mRNA sequence.
ACCESSION BU857520
VERSION BU857520.1 GI:24042512
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 927)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strauberg, Ph.D.
Email: rgs@remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM2887 row: k column: 22
High quality sequence stop: 667.
location/Qualifiers
1..927

FEATURES
SOURCE
1..927
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6647374"
/issue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_107"
/note="Organ: breast; Vector: pOTB7, Site 1: EcoRI;
Site 2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

ORIGIN

Query Match 18.4%; Score 805.8; DB 5; Length 927;
Best Local Similarity 97.5%; Pred. No. 7.3e-167;
Matches 841; Conservative 0; Mismatches 17; Indels 5; Gaps 2;
Qy 887 GGTAAAGAGATACAGAAAGAAAGGCAAGCTGTTTCAAGGCTATGAAACTGAAAG 946
Db 8 GGTAAAGAGATACAGAAAGAAAGGCAAGCTGTTTCAAGGCTATGAAACTGAAAG 67
Qy 947 GAGAGAGAAACAGAGCTATCTGAGAAATTAACCTGAGTGCAGCGGAGCTTTCCGA 1006
Db 68 GAGAGAGAAACAGAGCTATCTGAGAAATTAACCTGAGTGCAGCGGAGCTTTCCGA 127
Qy 1007 GACATCCAGACTCTGCTCCCAAGCCCTTCTCATGTCGGCGAGTGAAGGACATTA 1066
Db 128 GACATCCAGACTCTGCTCCCAAGCCCTTCTCATGTCGGCGAGTGAAGGACATTA 187
Qy 1067 AAGGAATCCCATTTGGAAGTACAGAAAGAAAGACTCTGTTTAAAGTGGTCTCTGA 1126
Db 188 AAGGAATCCCATTTGGAAGTACAGAAAGAAAGACTCTGTTTAAAGTGGTCTCTGA 247
Qy 1127 ATTTTCAAAAGTCAAAACAAAACTCCTAAGCACTCTCTATTAAAGGAAACCTGTGG 1186
Db 248 ATTTTCAAAAGTCAAAACAAAACTCCTAAGCACTCTCTATTAAAGGAAACCTGTGG 307
Qy 1187 TTCTTATCTGAAAATCTGTTTGAATTAAGAGGCAAGGCAAGGCAAGGCAAGGCA 1246
Db 308 TTCTTATCTGAAAATCTGTTTGAATTAAGAGGCAAGGCAAGGCAAGGCAAGGCA 367
Qy 1247 CCGGCTTCAAGTGAACACCCACCAAGACTCTCTCACTCCCAAAAGGAGCCAGACCA 1306
Db 368 CCGGCTTCAAGTGAACACCCACCAAGACTCTCTCACTCCCAAAAGGAGCCAGACCA 427
Qy 1307 TCCCAAGAGAAAGCCCTTCAAGTGAATTAAGGAAATTTCCGACCTTTCCACCAAGA 1366
Db 428 TCCCAAGAGAAAGCCCTTCAAGTGAATTAAGGAAATTTCCGACCTTTCCACCAAGA 487
Qy 1367 AATGATTTTGTGTCGTTGAGACCAAGCTCCCAATCAAGCTTCAATTAAGGAAATCTC 1426
Db 488 AATGATTTTGTGTCGTTGAGACCAAGCTCCCAATCAAGCTTCAATTAAGGAAATCTC 547
Qy 1427 TCCAAAGAGAGAGAGACTGTAGCAAGTGTCTGATCCATCAAGTGTTCAGAGAAAC 1486
Db 548 TCCAAAGAGAGAGAGACTGTAGCAAGTGTCTGATCCATCAAGTGTTCAGAGAAAC 607
Qy 1487 TTCAAGTCTTGGCTGTTCTCTTGAAGGACCACTCACTAGAGGCTTTAAGGAGCCCAA 1546
Db 608 TTCAAGTCTTGGCTGTTCTCTTGAAGGACCACTCACTAGAGGCTTTAAGGAGCCCAA 667
Qy 1547 TCCCTCAGACCTTTTGAAGAACTGATGACAGTGTGTTTTCGAAGGCGCATGAAAAC 1606
Db 668 TCCCTCAGACCTTTTGAAGAACTGATGACAGTGTGTTTTCGAAGGCGCATGAAAAC 727
Qy 1607 GAGAGTGTAGAGAGAGAGAGAAAGAGATGATTCAGAGATCAGGAAAC-AAAGAA 1665
Db 728 GAGAGTGTAGAGAGAGAGAGAGAAAGAGATGATTCAGAGATCAGGAAAC-AAAGAA 787
Qy 1666 TTTTACAGCACTGACCTGAGATGATTAATAAAGAAAGATTCAGAA---TCTGAG 1721
Db 788 TTTTACAGCACTGACCTGAGATGATTAATAAAGAAAGATTCAGAA---TCTGAG 847
Qy 1722 CTGAGGTTACCTCATTTTCCC 1744
Db 848 CTGAGGTTACCTCATTTTCCC 870

RESULT 8
BU517114 888 bp mRNA linear EST 12-SEP-2002
LOCUS AGENCOURT_10164432 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6514660
DEFINITION BU517114
ACCESSION BU517114
VERSION BU517114.1 GI:22824640

Query Match 17.6%; Score 770.8; DB 4; Length 817;
Best Local Similarity 98.3%; Pred. No. 3.9e-159;
Matches 800; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

QY 2597 CAAAATCTTTGAAAAGAAAGGAGGATGCTGTTGTTCAACAAGTTACAGCTGTAAAC 2656
DB 1 CAAAATCTTTGAAAAGAAAGGAGGATGCTGTTGTTCAACAAGTTACAGCTGTAAAC 60

QY 2657 AAAAGACTTATGATTTGGATGGCATGCAAAACCTGTATTAATGCTTGTATCAT 2716
DB 61 AAAAGACTTATGATTTGGATGGCATGCAAAACCTGTATTAATGCTTGTATCAT 120

QY 2717 CACTTCTCAAGTATTCCTTCATTTGGGCTTCATCTTTTACAGAACTCTTGTGTGGGA 2776
DB 121 CACTTCTCAAGTATTCCTTCATTTGGGCTTCATCTTTTACAGAACTCTTGTGTGGGA 180

QY 2777 TAGAGCTTAGAGGATGAGGAGAGAGTGTGAAAATAGTGTCTTCTTGGCTGGGAAA 2836
DB 181 TAGAGCTTAGAGGATGAGGAGAGAGTGTGAAAATAGTGTCTTCTTGGCTGGGAAA 240

QY 2837 TGTCTCATCTTAAACAAACAGATGATCTTAATGAGCTTCTCCATTCATCTTGTAAAA 2896
DB 241 TGTCTCATCTTAAACAAACAGATGATCTTAATGAGCTTCTCCATTCATCTTGTAAAA 300

QY 2897 TAAATTTGATGTATGATCTTGTGCTCTCCCTCCGTTTGTAAAAATATCAGGATA 2956
DB 301 TAAATTTGATGTATGATCTTGTGCTCTCCCTCCGTTTGTAAAAATATCAGGATA 360

QY 2957 GCACCTCCAGGCACTTTGGTCTCAGTGTAAATCCCTATTAACTATCTGAAAAGAAAAT 3016
DB 361 GCACCTCCAGGCACTTTGGTCTCAGTGTAAATCCCTATTAACTATCTGAAAAGAAAAT 420

QY 3017 AGAGCCAAAGACCTCTGGTCTCAAAATATATAGAAATGCTTCTTATAGTCTTCAGAGCTA 3076
DB 421 AGAGCCAAAGACCTCTGGTCTCAAAATATATAGAAATGCTTCTTATAGTCTTCAGAGCTA 480

QY 3077 TTGTGTGAAAACAAGTAGGGGCTTAATCTCTAGAAAGTAGGGGCTTTTATCTTAAAGA 3136
DB 481 TTGTGTGAAAACAAGTAGGGGCTTAATCTCTAGAAAGTAGGGGCTTTTATCTTAAAGA 540

QY 3137 GAATATGTCCTCCAGATTTATTAAGCATTTTAAAGAGAGCAAGCAAGTATGTAAGGCTGTG 3196
DB 541 GAATATGTCCTCCAGATTTATTAAGCATTTTAAAGAGAGCAAGCAAGTATGTAAGGCTGTG 600

QY 3197 GCGGGCCCAATCAGTGGAGCAAGAGAGAGATGGGATACATGTTGGGAGAGAGAGAAA 3256
DB 601 GCGGGCCCAATCAGTGGAGCAAGAGAGAGATGGGATACATGTTGGGAGAGAGAGAAA 660

QY 3257 GTTCTCAGGGGCTCCCACTGCTAAAGTTTGTGAGATGTGATCTGTGCTTCTCG 3316
DB 661 GTTCTCAGGGGCTCCCACTGCTAAAGTTTGTGAGATGTGATCTGTGCTTCTCG 718

QY 3317 ATTGACTTTTAAAGAAATTTCTGCGACACATGTAATTTCTTGATGATCTTGCTG 3376
DB 719 ATTGACTTTTAAAGAAATTTCTGCGACACATGTAATTTCTTGATGATCTTGCTG 778

QY 3377 CTCTTATTTCTCTTTTGTGTGTGTGTGTGTG 3410
DB 779 CTCTTATTTCTCTTTTGTGTGTGTGTGTGTG 812

RESULT 10
LOCUS BU929398 792 bp mRNA linear EST 18-OCT-2002
DEFINITION AGENCOURT_10425817 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:6663225
5', mRNA sequence.
ACCESSION BU929398
VERSION BU929398.1 GI:24118128
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 792)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLCM2928 row: P column: 09
High quality sequence stop: 554.
Location/Qualifiers

FEATURES
source
1..792
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6663225"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_79"
/note="Organ: Placenta; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgctcgccg); Site 2: SfiI (ggccatctatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCATTATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 17.5%; Score 766; DB 5; Length 792;
Best Local Similarity 98.0%; Pred. No. 4.5e-158;
Matches 775; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 3563 GATTTCTAGAAAACATGCGCCAGATTTCTTGTCTTCATCATTTTATGTCAGGCT 3622
DB 2 GATTTCTAGAAAACATGCGCCAGATTTCTTGTCTTCATCATTTTATGTCAGGCT 61

QY 3623 GTTCAGTTTCTTACCTTACCTATGATGATTTCTTGGTAAGTGTCCAAAAAGAAA 3682
DB 62 GTTCAGTTTCTTACCTTACCTATGATGATTTCTTGGTAAGTGTCCAAAAAGAAA 121

QY 3683 AGACCAATCAGTGTCTCTTGAATTTGTTGATCCCTCAGTTTCTTGTATTTGAG 3742
DB 122 AGACCAATCAGTGTCTCTTGAATTTGTTGATCCCTCAGTTTCTTGTATTTGAG 181

QY 3743 CATGTGTGGGTTCTTAATTTTGGTATGATGATGAGAAATTTAACCATTTGTGTGCC 3802
DB 182 CATGTGTGGGTTCTTAATTTTGGTATGATGATGAGAAATTTAACCATTTGTGTGCC 241

QY 3803 CTACCCAGGGAGATCCCCAGTTTCTGACTGAGTATGATGAGAGAAATCCAGAGGTGC 3862
DB 242 CTACCCAGGGAGATCCCCAGTTTCTGACTGAGTATGATGAGAGAAATCCAGAGGTGC 301

QY 3863 TATCTGCGCAGATTTATGATGATTTCTTGTGTTCTTCTCCCTCGAGGACCTCT 3922
DB 302 TATCTGCGCAGATTTATGATGATTTCTTGTGTTCTTCTCCCTCGAGGACCTCT 361

QY 3923 TATTTATTTGTCCTCTTCTTCTAGTTAATTTCTCTTTGATTTGATTTGAGAGAG 3982
DB 362 TATTTATTTGTCCTCTTCTTCTAGTTAATTTCTCTTTGATTTGATTTGAGAGAG 421

QY 3983 GTTGGACAGTATGATTTGAGAAAGTTTCCAGTGAAGTAAATTTACAGTGTATGAGTGGGG 4042
DB 422 GTTGGACAGTATGATTTGAGAAAGTTTCCAGTGAAGTAAATTTACAGTGTATGAGTGGGG 481

QY 4043 GAAATTAAGTCTTAATTTTCTTACATGAGGATCAACACTGTGAATTCATTTCAACTG 4102

Seq primer: M13 FORWARD
POLYA=yes.

FEATURES
source
Location/Qualifiers
1..765
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5833443"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_11b="NCI CGAP ED1"
/note="Organ: Left Pubic Bone; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP ED1 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma cell line C55. The library was constructed according to Bonaldo, Lemmon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is GCTCAAGGCT.
TAG_TISSUE=chondrosarcoma
TAG_LIB=UI-H-ED1
TAG_SEQ=CCTCAAGGCT"

ORIGIN

Query Match 17.3%; Score 756.2; DB 5; Length 765;
Best Local Similarity 99.5%; Pred. No. 6.5e-156;
Matches 758; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 946 AGAAGAGAGAAACAGACCTATCTGAGAAATTAATGAGTGCACCGAGCTTTCCG 1005
DB 765 AGAGAGAGAGAAACAGACCTATCTGAGAAATTAATGAGTGCACCGAGCTTTCCG 706
QY 1006 AGACATCCCGACACTGCGCTCCCAAGCCCTTCATCTGGGCGGAGTGGAGAAAGGACATA 1065
DB 705 AGACATCCCGACACTGCGCTCCCAAGCCCTTCATCTGGGCGGAGTGGAGAAAGGACATA 646
QY 1066 AAGAGAAATCCCATTTGGAAGTACAGAAAGAAAGCTCCTGTTAAAGAGGAACTGCTG 1125
DB 645 AAGAGAAATCCCATTTGGAAGTACAGAAAGAAAGCTCCTGTTAAAGAGGAACTGCTG 586
QY 1126 AATTTCAAAAGTCAAAACAAAACCTCTAAGCACTCTCTATTAAAGAGAACTGCTG 1185
DB 585 AATTTCAAAAGTCAAAACAAAACCTCTAAGCACTCTCTATTAAAGAGAACTGCTG 526
QY 1186 GTTCCTTATCTGAACTGTTTGAAGTGAATTTGAGAGCCAAAGAAACCCAGAAAAGC 1245
DB 525 GTTCCTTATCTGAACTGTTTGAAGTGAATTTGAGAGCCAAAGAAACCCAGAAAAGC 466
QY 1246 CCGGATCTTCAAGTGAACACCCCAACAGACTCTCCCAAAAGGAGCCAGACACC 1305
DB 465 CCGGATCTTCAAGTGAACACCCCAACAGACTCTCCCAAAAGGAGCCAGACACC 406
QY 1306 ATCCCAAGAGAAAGCCCTTCTCAAGTGAATGAATTTGCGTACCTTTCCACCAAG 1365
DB 405 ATCCCAAGAGAAAGCCCTTCTCAAGTGAATGAATTTGCGTACCTTTCCACCAAG 346
QY 1366 AATGATTTTGTGTCTTGGGACACGCTCCCATATCAAGTTTACGGAATCT 1425
DB 345 AATGATTTTGTGTCTTGGGACACGCTCCCATATCAAGTTTACGGAATCT 286
QY 1426 CTCGAAAGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1485
DB 285 CTCGAAAGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 226
QY 1486 CTTCAGTCTTGGCTTTCTTTTGAAGGAGCACTGAGTGAAGCTCTTAAGGAGCCAA 1545

DB 225 CTTCAGTCTTGGCTTTCTTTCTTTTGAAGGAGCACTGAGTGAAGCTCTTAAGGAGCCAA 166
QY 1546 ATCTTCAGACCTTTTGAAGAACCTGATGACAGTGTGTTTGAAGCGGCATGCAAAAC 1605
DB 165 ATCTTCAGACCTTTTGAAGAACCTGATGACAGTGTGTTTGAAGCGGCATGCAAAAC 106
QY 1606 TGGAGCTGATGAG 1665
DB 105 TGGAGCTGATGAG 46
QY 1666 TTTTACAGCAGCTGACGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1707
DB 45 TTTTACAGCAGCTGACGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 4

RESULT 13
BM475963
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM475963 1032 bp mRNA linear EST 05-FEB-2002
AGENCOURT_6480858 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5579934
EST. BM475963.1 GI:18525014
EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1032)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strauberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Plate: LLMK12357 row: g column: 07
High quality sequence stop: 653.

FEATURES
source

Location/Qualifiers
1..1032
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5579934"
/tissue_type="retinoblastoma"
/lab_host="NIH_MGC_67"
/clone_11b="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."

ORIGIN

Query Match 17.1%; Score 745.8; DB 4; Length 1032;
Best Local Similarity 94.4%; Pred. No. 1.4e-153;
Matches 852; Conservative 0; Mismatches 37; Indels 14; Gaps 7;

QY 1495 TGGCTGTTCTTGTGAGGAGACCACTGATGAGCTCTTAAGGAGCCAAATCCTTCAAG 1554
DB 36 TAGCAATTCCTTTTGAAGGAGACCACTGATGAGCTCTTAAGGAGCCAAATCCTTCAAG 95
QY 1555 ACCTTTGAAGAACTGATGACAGTGTGTTTGAAGCGGCATGCAAACTGAGAGCTG 1614
DB 96 ACCTTTGAAGAACTGATGACAGTGTGTTTGAAGCGGCATGCAAACTGAGAGCTG 155
QY 1615 ATGAG 1674
DB 156 ATGAG 215

```
QY 1675 GACTGAGCTCAGATATGATATTAAGAAAGAAATTCAGATCTGAGCTGAGCTTACCT 1734
DB 216 GACTGAGCTCAGATATGATATTAAGAAAGAAATTCAGATCTGAGCTGAGCTTACCT 275
QY 1735 CATTTCCTGAGCCAGATGATGTAAGAGTTGATGATTTACCCCTTCTGCTGTTG 1794
DB 276 CATTTCCTGAGCCAGATGATGTAAGAGTTGATGATTTACCCCTTCTGCTGTTG 335
QY 1795 TAGCATTTGGAGACACCATTAACCAAAATTAATCCACAGAAATTTTGGCTTACCTGTTGG 1854
DB 336 TAGCATTTGGAGACACCATTAACCAAAATTAATCCACAGAAATTTTGGCTTACCTGTTGG 395
QY 1855 ATAGAGGTCAGCCAGATGATGTAAGAGTTGATGATTTACCCCTTCTGCTGTTG 1914
DB 396 ATAGAGGTCAGCCAGATGATGTAAGAGTTGATGATTTACCCCTTCTGCTGTTG 455
QY 1915 GGAATATGCTGCTGCTGCAAGAAACCCCTGCTTCCAGATGTTGATGATGCTTCCGAG 1974
DB 456 GGAATATGCTGCTGCTGCAAGAAACCCCTGCTTCCAGATGTTGATGATGCTTCCGAG 515
QY 1975 AGTGGCAGAGACCTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 2034
DB 516 AGTGGCAGAGACCTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 575
QY 2035 ACCCTTCATACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2094
DB 576 ACCCTTCATACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 635
QY 2095 TCGGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2153
DB 636 TCGGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 695
QY 2154 GGGCTTCATACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2212
DB 696 GGGCTTCATACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 755
QY 2213 ACA-TTGGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2269
DB 756 ACA-TTGGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 815
QY 2270 AACGAGCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2327
DB 816 AACGAGCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 875
QY 2328 CTCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2380
DB 876 CTCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 935
QY 2381 ACC 2383
DB 936 CCC 938
```

RESULT 14
BO898842 891 bp mRNA linear EST 16-AUG-2002
LOCUS BO898842
DEFINITION AGENCOURT 8121821 lupski dorsal root ganglion Homo sapiens cDNA
clone IMAGE:6181536 5', mRNA sequence.
ACCESSION BO898842
VERSION BO898842.1 GI:22220856
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 891)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strauberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LLM43566 row: b column: 01
High quality sequence strip: 653.

FEATURES
Location/Qualifiers
1. 891
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6181536"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/clone_lib="lupski dorsal root ganglion"
/note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGCG-3' and
5'-GACTAGCTTCAGATCGCAGCGGCCGCTT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."

ORIGIN
Query Match 17.0%; Score 744.6; DB 5; Length 891;
Best Local Similarity 98.6%; Fred. No. 2.5e-153;
Matches 771; Conservative 0; Mismatches 7; Indels 4; Gaps 2;

```
QY 2834 AAATGCTCATCTTGAACAAACAGATGATGATGATGATGATGATGATGATGATGAT 2893
DB 2 AAATGCTCATCTTGAACAAACAGATGATGATGATGATGATGATGATGATGATGAT 61
QY 2894 AAATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2953
DB 62 AAATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121
QY 2954 ATAGCACTCCGAGCCACTTGGTCTCAGTGAAGATCCCTATTACTATCTGAAGAA 3013
DB 122 ATAGCACTCCGAGCCACTTGGTCTCAGTGAAGATCCCTATTACTATCTGAAGAA 181
QY 3014 AATAGAGCAAGACCTGCTGCTCAATATATAGAAATGCTTCTTATAGTCTGAGGA 3073
DB 182 AATAGAGCAAGACCTGCTGCTCAATATATAGAAATGCTTCTTATAGTCTGAGGA 241
QY 3074 CTATGCTGAAAAACAAGTAGGGGCTTAATCTCTTGAAGAGTAGGGCTTTATCTTAA 3133
DB 242 CTATGCTGAAAAACAAGTAGGGGCTTAATCTCTTGAAGAGTAGGGGCTTTATCTTAA 301
QY 3134 AAGAAATATGTCCTCCAGATTTATGACATTTTGAAGAGAGCAAGATGATGAGGTGT 3193
DB 302 AAGAAATATGTCCTCCAGATTTATGACATTTTGAAGAGAGCAAGATGATGAGGTGT 361
QY 3194 GTGGCTGGCCCATCACTGAGAGCAAGAGAGAGAGATGAGATCACTGTGGGAAGAGA 3253
DB 362 GTGGCTGGCCCATCACTGAGAGCAAGAGAGAGAGATGAGATCACTGTGGGAAGAGA 421
QY 3254 AAAGTTCCTCAGGGGCTCCCACTGCTAAAGTTTTTGTGAGATGTTGATCTGTGCTTC 3313
DB 422 AAAGTTCCTCAGGGGCTCCCACTGCTAAAGTTTTTGTGAGATGTTGATCTGTGCTTC 481
QY 3314 TGAATTTGACTTTTAAAGAAATATTTCTGACAGACATGATGATTTCTGATGATCTTG 3373
DB 482 TGAATTTGACTTTTAAAGAAATATTTCTGACAGACATGATGATTTCTGATGATCTTG 541
QY 3374 CTGCTCTTATTTCTCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 3433
DB 542 CTGCTCTTATTTCTCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 601
```

QY 3434 GTAACCTCATCTCTAGAGAGAGCTGCTCTATAGAGGAACTGCTGTAACTTCATT 3493
DB 602 GTAACTCCATCTCTAGAGAGAGCTGCTCTATAGAGGAACTGCTGTAACTTCATT 661
QY 3494 GCAGCAAGATGTAGAGAGAAATAGACTTAATTCACCTAGGGCTCTCATTCACACCT 3553
DB 662 GCAGCAAGATGTAGAGAGAAATAGACTTAATTCACCTAGGGCTCTCATTCACACCT 721
QY 3554 TAAGAGAGATTTCTAGAAAAA--CTGGGCAAGATTTCTTT--GTTCCATCATTTT 3609
DB 722 TAAGAGAGATTTCTAGAAAAAAGCTGGGCAAGATTTCTTTGTTCTTCATCATTTT 781
QY 3610 AA 3611
DB 782 AA 783

RESULT 15
BQ948768 931 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT 8805024 lupski_sciatic_nerve Homo sapiens cDNA clone
DEFINITION IMAGE:6197047 5', mRNA sequence.
ACCESSION BQ948768
VERSION BQ948768.1 GI:22364246
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 931)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM13606 row: h column: 08
High quality sequence stop: 640.
Location/Qualifiers
1. 931
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6197047"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_host="DH10B"
/clone_1ib="Lupski_sciatic_nerve"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:
NotI; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCG-3' and
5'-GACTAGTTCATGACGAGCGGCGGCGCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."

ORIGIN

Query Match 16.9%; Score 737.4; DB 5; Length 931;
Best Local Similarity 95.8%; Pred. NO. 9.6e-152;
Matches 857; Conservative 0; Mismatches 21; Indels 17; Gaps 9;

QY 3346 GCACATGTAGATTTCTTGATGATCTTGCTGCTCTATTCTCTTTGTGTGTGTGT 3405

DB 1 GCACATGTAGATTTCTTGATGATCTTGCTGCTCTATTCTCTTTGTGTGTGTGT 60
QY 3406 GT 3465
DB 61 GT 120
QY 3466 TATAGAGAACTGCTGTAACTTCATGTGAGAGAGATGTAGAGAAATAGACTTAA 3525
DB 121 TATAGAGAACTGCTGTAACTTCATGTGAGAGAGATGTAGAGAAATAGACTTAA 180
QY 3526 TTCCACTAGGGGCTTCATCTCACACCTTAAAGAGAGATTTCTAGAAAACTGGGCGAG 3585
DB 181 TTCCACTAGGGGCTTCATCTCACACCTTAAAGAGAGATTTCTAGAAAACTGGGCGAG 240
QY 3586 ATTTCTTTGTTCTTCATCATTTTAAATGAGGAGGCTGTCAGTTTCTTACTCTTACCT 3645
DB 241 ATTTCTTTGTTCTTCATCATTTTAAATGAGGAGGCTGTCAGTTTCTTACTCTTACCT 300
QY 3646 ATGTGATATTTCTTGTAACTGTCACAAAAAGAAAAAGAACCCGATCAGTGTCTTGAC 3705
DB 301 ATGTGATATTTCTTGTAACTGTCACAAAAAGAAAAAGAACCCGATCAGTGTCTTGAC 360
QY 3706 TTTGTTCTTTGATCCCTCACTTCTTCTGATTTTCAAGATGTGTGGGTTCTTAATTTTG 3765
DB 361 TTTGTTCTTTGATCCCTCACTTCTTCTGATTTTCAAGATGTGTGGGTTCTTAATTTTG 420
QY 3766 GGTATAGTATGAAATTTTAAACATGTTGTGTCCTTCCGACGAGGAGATCCCGAGTTT 3825
DB 421 GGTATAGTATGAAATTTTAAACATGTTGTGTCCTTCCGACGAGGAGATCCCGAGTTT 480
QY 3826 CTGACTTGAAGTATGACTGAGAGAAATCCAGAGTGTCTATCTGAGCATTTAATAGAT 3885
DB 481 CTGACTTGAAGTATGACTGAGAGAAATCCAGAGTGTCTATCTGAGCATTTAATAGAT 540
QY 3886 TCTATTTCTTGTGTTCTCCCTCTCCCTGAGGACCTTATTTTATGTGCTCTCTAG 3945
DB 541 TCTATTTCTTGTGTTCTCCCTCTCCCTGAGGACCTTATTTTATGTGCTCTCTAG 600
QY 3946 GTTAATTTCTTGTATTTGATCTTTGTGAGAGAGTGTGACAGATGATTTGCAAGT 4005
DB 601 GTTAATTTCTTGTATTTGATCTTTGTGAGAGAGTGTGACAGATGATTTGCAAGT 660
QY 4006 TCCAGTGCAGAAATTAAGTGTGTAGAGTGTGGGGG--AAATTAAGTCTTA--TTTTT 4062
DB 661 TCCAGTGCAGAAATTAAGTGTGTAGAGTGTGGGGGAAATTAAGTCTTAATTTTTC 720
QY 4063 CCTACATGGG-ATACAACT-GTGAATTCATCTTCACTG-AAAGCCCTGAGTTCTC 4119
DB 721 CCTACATGGGATACAACTGATGATTCATCTTCACTGAAAGCCCTGAGTTCTC 780
QY 4120 CTAAACATP---GTTGTTGTTTCTTTTAACTAA--GTTAAGTACTGTTAATAA 4173
DB 781 TTAACCTTAAGTGTGGTGTGTTTCTTTTAACTAAAGTTTAAGTGTAGGTTAATAA 840
QY 4174 TTAACCTTAAGTGTGGTGTGTTTCTTTTAACTAAAGTTTAAGTGTAGGTTAATAA 8423
DB 841 TTAACCTTAAGTGTGGTGTGTTTCTTTTAACTAAAGTTTAAGTGTAGGTTAATAA 895

Search completed: March 24, 2005, 20:33:07
Job time: 13109 secs

This Page Blank (uspto)

Dh 716 AAGATGGGATATTACAGAGATCAGGGAACAAGATTTTACAGGACTGCAAGCTCAGAAT 775
Qy 1691 GTTAAAGAAAGAAAGAAATTCAGGAATCTGAGCTGAGCTTACCTCATTTTCCCTGAGCC 1750
Dh 776 GTATAAAGAAAGAAAGAAATTCAGGAATCTGAGCTGAGCTTACCTCATTTTCCCTGAGCC 835
Qy 1751 AGATGATGTTGAAGTTTGATGATTAACCCCTTCCTGCTGTTGATGATTTGGAGAGACC 1810
Dh 836 AGATGATGTTGAAGTTTGATGATTAACCCCTTCCTGCTGTTGATGATTTGGAGAGACC 895
Qy 1811 ATTACCAAAATTAATCTCACAAGAAATTTGAGCTAACCCCTGATGAGATGAGCCGATG 1870
Dh 896 ATTACCAAAATTAATCTCACAAGAAATTTGAGCTAACCCCTGATGAGATGAGCCGATG 955
Qy 1871 CAGATTGAGATCCAGAAGAGCAACACTCACCGGAGCTGTGAGAAATAGCTGTGCTG 1930
Dh 956 CAGATTGAGATCCAGAAGAGCAACACTCACCGGAGCTGTGAGAAATAGCTGTGCTG 1015
Qy 1931 GCAGAGAACCTGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1990
Dh 1016 GCAGAGAACCTGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1075
Qy 1991 ATATGACCTTTGCTCTCACAATATGTTATCACTCGCTGATTAATACCTTTCACTACTTCC 2050
Dh 1076 ATATGACCTTTGCTCTCACAATATGTTATCACTCGCTGATTAATACCTTTCACTACTTCC 1135
Qy 2051 TTGACTTTGTTTTCATTACTCTGTGATTTACAAAACCTCTTTCATTCGGGCTAAATTTGAGT 2110
Dh 1136 TTGACTTTGTTTTCATTACTCTGTGATTTACAAAACCTCTTTCATTCGGGCTAAATTTGAGT 1195
Qy 2111 TATGAGGGGATTTGGGATTTCTTTCCCTTTTGGGAAATGGGCTCTCAAGCTAAAGC 2170
Dh 1196 TATGAGGGGATTTGGGATTTCTTTCCCTTTTGGGAAATGGGCTCTCAAGCTAAAGC 1255
Qy 2171 TATAGATGCGCAGATTCAGAAGTTTCAGGGGCTGTTTCTATACATTTGCTATGTTAAA 2230
Dh 1256 TATAGATGCGCAGATTCAGAAGTTTCAGGGGCTGTTTCTATACATTTGCTATGTTAAA 1315
Qy 2231 GGGGTAAAGGGCTCTCTTCATTAGACATGTGGAAGATGAAGCAAGCCCTTCTTTAGAG 2290
Dh 1316 GGGGTAAAGGGCTCTCTTCATTAGACATGTGGAAGATGAAGCAAGCCCTTCTTTAGAG 1375
Qy 2291 CTGTGCTGATGGCACTCTTCTCACCCCTGTACACCCCTCTTAATGCTGGTATAGAT 2350
Dh 1376 CTGTGCTGATGGCACTCTTCTCACCCCTGTACACCCCTCTTAATGCTGGTATAGAT 1435
Qy 2351 TTTTAAACCTTAATTAATAAACAACAACCTCACCATGACTTTAGAGACCAAGAGATG 2410
Dh 1436 TTTTAAACCTTAATTAATAAACAACAACCTCACCATGACTTTAGAGACCAAGAGATG 1495
Qy 2411 ACAAGTGAAGCGATGAAGCAAGCCATCTTCAAGAGTGAAGAAAGACATCGGAGATTGAT 2470
Dh 1496 ACAAGTGAAGCGATGAAGCAAGCCATCTTCAAGAGTGAAGAAAGACATCGGAGATTGAT 1555
Qy 2471 AGATACTGCTGTAAGAAAGATAGTTGTTCAATTTGAGAAATATTTCTGTGATTAAGTCAATG 2530
Dh 1556 AGATACTGCTGTAAGAAAGATAGTTGTTCAATTTGAGAAATATTTCTGTGATTAAGTCAATG 1615
Qy 2531 GAAGGATGTTTGGCTGTGATTAATTTTTCAGTTAATGATAACAATTTCTTTACTGCTC 2590
Dh 1616 GAAGGATGTTTGGCTGTGATTAATTTTTCAGTTAATGATAACAATTTCTTTACTGCTC 1675
Qy 2591 AAAAACCAAAATCTTTGGAAAAGAAAGTGGGATGTTAGTTTTCAGAAACAAGTTACAGCT 2650
Dh 1676 AAAAACCAAAATCTTTGGAAAAGAAAGTGGGATGTTAGTTTTCAGAAACAAGTTACAGCT 1735
Qy 2651 GTTAAACAAAAGCACTTAGTATTTGGGATGGCATGCCAAAACCTGTATAAATGTCCTTGTA 2710
Dh 1736 GTTAAACAAAAGCACTTAGTATTTGGGATGGCATGCCAAAACCTGTATAAATGTCCTTGTA 1795
Qy 2711 TCACATCACTTCAAGATATTTCTTCAATTGGGCTTCAATCCCTTTTGAAGAGAACTCTGGTG 2770
Dh 1796 TCACATCACTTCAAGATATTTCTTCAATTGGGCTTCAATCCCTTTTGAAGAGAACTCTGGTG 1855

Qy 2771 GTGGGATAGAGACTTTAGGGAGGGTATAGGGGAGAGTGTGAAAATAGGTCCTTCTTGGCT 2830
Dh 1856 GTGGGATAGAGACTTTAGGGAGGGTATAGGGGAGAGTGTGAAAATAGGTCCTTCTTGGCT 1915
Qy 2831 GGCATAATGCTACATCTTGAACCAACAGATGATCACTTAATGAGCTTCTCAATCACTTTG 2890
Dh 1916 GGCATAATGCTACATCTTGAACCAACAGATGATCACTTAATGAGCTTCTCAATCACTTTG 1975
Qy 2891 TAAAAATTAATTTGATATGATACCATCTTGGTCTCTCCCTCCGTTTGTAAAAATATC 2950
Dh 1976 TAAAAATTAATTTGATATGATACCATCTTGGTCTCTCCCTCCGTTTGTAAAAATATC 2035
Qy 2951 AGGATAGCACTCCAGGCCACTTTGGTCTCAGTGAAGATTCCTATTAATCAATCTGAAG 3010
Dh 2036 AGGATAGCACTCCAGGCCACTTTGGTCTCAGTGAAGATTCCTATTAATCAATCTGAAG 2095
Qy 3011 GAAATATAGAGCCAAAGCTCTGGTCTCAAAATATATAGAAATTCCTTTTATGCTTCA 3070
Dh 2096 GAAATATAGAGCCAAAGCTCTGGTCTCAAAATATATAGAAATTCCTTTTATGCTTCA 2155
Qy 3071 GACATAATTTGTGAAGAAACAAGTATGGGCTTAATCTCTGAAAGTATGGGCTTTATCTT 3130
Dh 2156 GACATAATTTGTGAAGAAACAAGTATGGGCTTAATCTCTGAAAGTATGGGCTTTATCTT 2215
Qy 3131 TAAAGGAATATATGTCCTCCAGATTAATTAAGCACTTTTGAAGAAAGCCAAAGTATGTAGG 3190
Dh 2216 TAAAGGAATATATGTCCTCCAGATTAATTAAGCACTTTTGAAGAAAGCCAAAGTATGTAGG 2275
Qy 3191 TGTGTGGCTGGGCCATCAGTGGAGCAGAGAGAAATGGGAATACCATTTGGGAAAGAG 3250
Dh 2276 TGTGTGGCTGGGCCATCAGTGGAGCAGAGAGAAATGGGAATACCATTTGGGAAAGAG 2335
Qy 3251 AGAAAAGTTCCTCAGGGGCTCCCACTGCTAAAGTTTTTGTGAGATGTTGATCTGTGCT 3310
Dh 2336 AGAAAAGTTCCTCAGGGGCTCCCACTGCTAAAGTTTTTGTGAGATGTTGATCTGTGCT 2395
Qy 3311 TCTGTGATTTGACTTTTAAAGGAATTAATCTGGCAGACATGTAGATTTCTTGGATGATC 3370
Dh 2396 TCTGTGATTTGACTTTTAAAGGAATTAATCTGGCAGACATGTAGATTTCTTGGATGATC 2455
Qy 3371 TTGCTGCTTAATTTCTCCCTTTTGTGATGATGATGATGATGATGATGATGATGATGATG 3430
Dh 2456 TTGCTGCTTAATTTCTCCCTTTTGTGATGATGATGATGATGATGATGATGATGATGATG 2515
Qy 3431 TTTGTAATCTCCATCTGCTTAGAGAGTGGGCTCTCTATAAGGGAACCTGCTGTAACTTC 3490
Dh 2516 TTTGTAATCTCCATCTGCTTAGAGAGTGGGCTCTCTATAAGGGAACCTGCTGTAACTTC 2575
Qy 3491 ATTGCAAGCAAGATGTAGAGAGAAATAGACTTAATTTCCATAGGGGCTCTCATCTCACA 3550
Dh 2576 ATTGCAAGCAAGATGTAGAGAGAAATAGACTTAATTTCCATAGGGGCTCTCATCTCACA 2635
Qy 3551 CCTTAAGGAGGAATTTCTAGAAAACCTGGGCCAGATTTTCTTGTCTCCATCATTTTA 3610
Dh 2636 CCTTAAGGAGGAATTTCTAGAAAACCTGGGCCAGATTTTCTTGTCTCCATCATTTTA 2695
Qy 3611 ATGTGCAAGGCTGTACAGTTTCTTACTCTTACCTATGTGATATTTCTTGTGATACGTC 3670
Dh 2696 ATGTGCAAGGCTGTACAGTTTCTTACTCTTACCTATGTGATATTTCTTGTGATACGTC 2755
Qy 3671 GAAAAAGAAAAAGACCAATCAGTGTCTCTTGACTTTGTTCTTGTATCCCTCAGTTTCT 3730
Dh 2756 GAAAAAGAAAAAGACCAATCAGTGTCTCTTGACTTTGTTCTTGTATCCCTCAGTTTCT 2815
Qy 3731 TCTTGAATTTCAGATGTGTGCGGTTCTCTAATTTTGGGATATGAGTATGCAAAATTTAACAT 3790
Dh 2816 TCTTGAATTTCAGATGTGTGCGGTTCTCTAATTTTGGGATATGAGTATGCAAAATTTAACAT 2875
Qy 3791 TGTGTTTGGGCCCTTACCCAGGGGACTCCCAAGTTTCTGACTTGAAGTGAAGCTGAGAAGAA 3850
Dh 2876 TGTGTTTGGGCCCTTACCCAGGGGACTCCCAAGTTTCTGACTTGAAGTGAAGCTGAGAAGAA 2935

QY 3851 TCACAGAGGTGCTATCTGGCCAGATTAAAGTAGATTCTATTTCTGTTCTCCCTCTCC 3910
DB 2936 TCACACAGAGGTGCTATCTGGCCAGATTAAAGTAGATTCTATTTCTGTTCTCCCTCTCC 2995
QY 3911 CTGAGGACCGCTATTTATTTGTCCTCTTCAAGTTAATTCCTGTTAGTTGACTTT 3970
DB 2996 CTGAGGACCGCTATTTATTTATTTGTCCTCTTCAAGTTAATTCCTGTTAGTTGACTTT 3055
QY 3971 GTTGAGAGAGGTTGAGCAGTAGATTAGCAAGTTCCAAAGTTTACAGTGTGT 4030
DB 3056 GTTGAGAGAGGTTGAGCAGTAGATTAGCAAGTTCCAAAGTTTACAGTGTGT 3115
QY 4031 AGAGTGTGGGGGAAAATTAGTCTTATTTTCCCTACATGGATPACAACCTGTGAATTC 4090
DB 3116 AGAGTGTGGGGGAAAATTAGTCTTATTTTCCCTACATGGATPACAACCTGTGAATTC 3175
QY 4091 AATCTTCACTGAGAGCCCTGAGTCTCTCTTAAACATAGTTGTTGTTTCTTTTAAACA 4150
DB 3176 AATCTTCACTGAGAGCCCTGAGTCTCTCTTAAACATAGTTGTTGTTTCTTTTAAACA 3235
QY 4151 AAGTTTAACTGATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4210
DB 3236 AAGTTTAACTGATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3295
QY 4211 TTTTATGCCATTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 4270
DB 3296 TTTTATGCCATTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3355
QY 4271 AATGTTAAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 4330
DB 3356 AATGTTAAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3415
QY 4331 TGTGCTGTT 4340
DB 3416 TGTGCTGTT 3425

RESULT 2
AK125654 3425 bp mRNA linear PRI 09-SEP-2003
LOCUS Homo sapiens cDNA FLJ3666 fis, clone SYN04007012.
DEFINITION AK125654
ACCESSION AK125654
VERSION AK125654.1 GI.34531819
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Oca, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Ishii, S.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
Yamanoto, J., Isono, Y., Kawai, H., Saito, K., Nishikawa, T.,
Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
Kikuchi, H., Kanda, K., Waga, S., Muraoka, K., Kakehori, K.,
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,
Sugano, S., Nagahara, K., Masuno, Y., Nagai, K., and Iisagoi, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3425)
REFERENCE
AUTHORS Iisagoi, T. and Yamamoto, J.
TITLE Direct Submision
JOURNAL Submitted (15-JUL-2003) Takao Iisagoi, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0819, Japan
(E-mail: iisagoi@kazusa.or.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'-3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
COMMENT
Location/Qualifiers

FEATURES

source 1. 3425
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SYNOV4007012"
/tissue_type="synovial membrane tissue from rheumatoid
arthritis"
/clone_id="SYNOV4"
/note="Cloning vector: pME18SFL3"

ORIGIN
Query Match 66.1%; Score 2888.4; DB 9; Length 3425;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2889; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1451 AAGGTGCTGATGCTCAGTCAAGTTGTCAGAGAACTTCAAGTCTGCTGCTCTTCTT 1510
DB 536 AGGAGTCTGATGCTCAGTCAAGTTGTCAGAGAACTTCAAGTCTGCTGCTCTTCTT 595
QY 1511 GAGGAGCACTCAGTGAAGCTCTTAAGGAGCCAAATCTTTCAGACTTTTGAAGAACT 1570
DB 596 GAGGAGCACTCAGTGAAGCTCTTAAGGAGCCAAATCTTTCAGACTTTTGAAGAACT 655
QY 1571 GGATGACAGTGTGTTTTCAGAGCGGATGCAAACTGAGCTGATGAGAGAGAGAA 1630
DB 656 GGATGACAGTGTGTTTTCAGAGCGGATGCAAACTGAGCTGATGAGAGAGAGAA 715
QY 1631 AAGATGGATATTCAGAGATTCAGGAGCAAAATTTTACAGGAGCTGAGCTCAAGAT 1690
DB 716 AAGATGGATATTCAGAGATTCAGGAGCAAAATTTTACAGGAGCTGAGCTCAAGAT 775
QY 1691 GTATTAAGAAAGAAAGAAATTCAGAGATTCAGGAGCTGAGCTCAAGATTTTCCCTGAGCC 1750
DB 776 GTATTAAGAAAGAAAGAAATTCAGAGATTCAGGAGCTGAGCTCAAGATTTTCCCTGAGCC 835
QY 1751 AGATGATGTTGAAGTTGATGATTAACCCCTTCTGCTGTTGATGATTTGAGAGACC 1810
DB 836 AGATGATGTTGAAGTTGATGATTAACCCCTTCTGCTGTTGATGATTTGAGAGACC 895
QY 1811 AATACCAAAATTAATTAATTCAGAGATTTTACAGTCAACCTGTTGAGAGAGTACCGATG 1870
DB 896 AATACCAAAATTAATTAATTCAGAGATTTTACAGTCAACCTGTTGAGAGAGTACCGATG 955
QY 1871 CAGATTGAGATTCAGAGATTCAGAGATTCAGGAGCTGAGCTGAGTAAATAGCTGTGCTG 1930
DB 956 CAGATTGAGATTCAGAGATTCAGAGATTCAGGAGCTGAGCTGAGTAAATAGCTGTGCTG 1015
QY 1931 GCAGAAACCTGCTTCAATGATGTTGATGATTCATTCGAGAGTGGAGAGACCTGT 1990
DB 1016 GCAGAAACCTGCTTCAATGATGTTGATGATTCATTCGAGAGTGGAGAGACCTGT 1075
QY 1991 ATATGAGCTTGTGCTTCAATGATGTTGATGATTCATTCGAGAGTGGAGAGACCTGT 2050
DB 1076 ATATGAGCTTGTGCTTCAATGATGTTGATGATTCATTCGAGAGTGGAGAGACCTGT 1135
QY 2051 TTGACTTGTGTTTCAATGATGTTGATGATTCATTCGAGAGTGGAGAGACCTGT 2110
DB 1136 TTGACTTGTGTTTCAATGATGTTGATGATTCATTCGAGAGTGGAGAGACCTGT 1195
QY 2111 TATGAGAGGATGATGATGATTTCTTTTCCCTTTTGGAGAAATGGGCTTCAAGCTAAAGC 2170
DB 1196 TATGAGAGGATGATGATGATTTCTTTTCCCTTTTGGAGAAATGGGCTTCAAGCTAAAGC 1255
QY 2171 TATGAGAGGATGATGATGATTTCTTTTCCCTTTTGGAGAAATGGGCTTCAAGCTAAAGC 2230
DB 1256 TATGAGAGGATGATGATGATTTCTTTTCCCTTTTGGAGAAATGGGCTTCAAGCTAAAGC 1315
QY 2231 GGGGTAAAGGGGCTCTTTCATTAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2290
DB 1316 GGGGTAAAGGGGCTCTTTCATTAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1375
QY 2291 CTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2350

Db 1376 CTGTGCTGATGACACTCTTCTCACCCCTGTGACACCTCTTAATAGTGGGTAATAGTAT 1435
Qy 2351 TTTTAACCTTAATAATAAACAACAACCTCACCTAGCTTTAGAGCCGAAGAAGAAG 2410
Db 1436 TTTTAACCTTAATAATAAACAACAACCTCACCTAGCTTTAGAGCCGAAGAAGAAG 1495
Qy 2411 ACAAGTGAAGCGATGAAGCAAGCATCTTCAAGATAGAAAAAGATCGAGAGTGGGT 2470
Db 1436 ACAAGTGAAGCGATGAAGCAAGCATCTTCAAGATAGAAAAAGATCGAGAGTGGGT 1555
Qy 2471 AGATTAACGTCTGAAAAAGATAGTGTTCATTGAAACTAATCTGTGATACAGTCATGTGG 2530
Db 1556 AGATTAACGTCTGAAAAAGATAGTGTTCATTGAAACTAATCTGTGATACAGTCATGTGG 1615
Qy 2531 GAAGGAGATGTTGGCTGTGATTAATTTTTCAGTTAATGATAACAATTTCTTACTGCTC 2590
Db 1616 GAAGGAGATGTTGGCTGTGATTAATTTTTCAGTTAATGATAACAATTTCTTACTGCTC 1675
Qy 2591 AAAAAACAAATCTTTGAAAAAGAAAGTGGGATGTTAGTTACAGAAAGTTACAGCT 2650
Db 1676 AAAAAACAAATCTTTGAAAAAGAAAGTGGGATGTTAGTTACAGAAAGTTACAGCT 1735
Qy 2651 GTAAACAAAAGCACTTAGTATTGTTGGATGSCATGCAAAAACCTGTATTAATGTCCTGTGA 2710
Db 1736 GTAAACAAAAGCACTTAGTATTGTTGGATGSCATGCAAAAACCTGTATTAATGTCCTGTGA 1795
Qy 2711 TCACATCACTTCTCAAGTATTCCTTCATTTGGCTTATCTCTTTTAGACAACCTTTGGTG 2770
Db 1796 TCACATCACTTCTCAAGTATTCCTTCATTTGGCTTATCTCTTTTAGACAACCTTTGGTG 1855
Qy 2771 GTGGGATAGAGACTTAGGAGGGGTAGGGGAGAGTGTGAAATAGGTGCTCCCTTGGGCT 2830
Db 1856 GTGGGATAGAGACTTAGGAGGGGTAGGGGAGAGTGTGAAATAGGTGCTCCCTTGGGCT 1915
Qy 2831 GGCAATGTCATCATCTTGAAAACAACAAGATGACTTAATGAGCTTCTCATTCACCTTGG 2890
Db 1916 GGCAATGTCATCATCTTGAAAACAACAAGATGACTTAATGAGCTTCTCATTCACCTTGG 1975
Qy 2891 TTAATAATTAATTTGTATGTGTACCATTTGGTCTCTCCCTCCGTTTGTGTTAAATATC 2950
Db 1976 TTAATAATTAATTTGTATGTGTGTACCATTTGGTCTCTCCCTCCGTTTGTGTTAAATATC 2035
Qy 2951 AGATAGCACTCCCAAGGCACTTGGTCTCAGTGTAGATCCCTATTAATCTGTAAG 3010
Db 2036 AGATAGCACTCCCAAGGCACTTGGTCTCAGTGTAGATCCCTATTAATCTGTAAG 2095
Qy 3011 GAAAAATAGCAAGCACTCTGTGTCTCAATATATATAGAAATGCTTTCTTTAGTCTCA 3070
Db 2096 GAAAAATAGCAAGCACTCTGTGTCTCAATATATATAGAAATGCTTTCTTTAGTCTCA 2155
Qy 3071 GGACTATTTGTGTGAAAAACAAGTAGGGGCTTAATCTCTAGAAAGTGGGGCTTTTATCT 3130
Db 2156 GGACTATTTGTGTGAAAAACAAGTAGGGGCTTAATCTCTAGAAAGTGGGGCTTTTATCT 2215
Qy 3131 TTAAGAGATATGTCCCGAGATTATAGACATTTTAGAGAGAAAGCAAGATATGTAGGG 3190
Db 2216 TTAAGAGATATGTCCCGAGATTATAGACATTTTAGAGAGAAAGCAAGATATGTAGGG 3275
Qy 3191 TGTGTGCTGGCCATCAGTGAAGCAAGAAAGAGATGGGATACATTTGGGGAAGAGA 3250
Db 2276 TGTGTGCTGGCCATCAGTGAAGCAAGAAAGAGATGGGATACATTTGGGGAAGAGA 2335
Qy 3251 AGAAAAAGTTCCTAGGGGCTCCCACTGCTAAAGTTTTTGTGAGATGTTGATCTGTGCT 3310
Db 2336 AGAAAAAGTTCCTAGGGGCTCCCACTGCTAAAGTTTTTGTGAGATGTTGATCTGTGCT 2395
Qy 3311 TCCTGATTTGACTTTTAAGAAATTAATCTGGACACACATGTAGATCTTGGATGATC 3370
Db 2396 TCCTGATTTGACTTTTAAGAAATTAATCTGGACACACATGTAGATCTTGGATGATC 2455
Qy 3371 TTGCTGCTTATATTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3430
Db 2456 TTGCTGCTTATATTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2515

Qy 3431 TTTGTATACACATCTGCTTAGAGAGTGGGCTCTCATTAAGGGAACCTGCTGTAACCTC 3490
Db 2516 TTTGTATACCTCATCTGCTTAGAGAGTGGGCTCTCATTAAGGGAACCTGCTGTAACCTC 2575
Qy 3491 ATTGACAGAGATGTAGAGAAATAGGACTTAATTCCTACATAGGGGCTCTCATCTACA 3550
Db 2576 ATTGACAGAGATGTAGAGAAATAGGACTTAATTCCTACATAGGGGCTCTCATCTACA 2635
Qy 3551 CCTTAAGAGAGATTTCTAGAAAAAATGGGCAAGATTTTCTTGTCTCCATCATTTTA 3610
Db 2636 CCTTAAGAGAGATTTCTAGAAAAAATGGGCAAGATTTTCTTGTCTCCATCATTTTA 2695
Qy 3611 ATGTGGCAGGCTGTTCAAGTTTCTTACTCTTACCTTACATATGATATTTCTTGTAAGTGTG 3670
Db 2696 ATGTGGCAGGCTGTTCAAGTTTCTTACTCTTACATATGATATTTCTTGTAAGTGTG 2755
Qy 3671 CAAAAAGAAAAAAGCCCAATCAGTGTCTCTGACCTTGTCTTGTGATCCCTCAGTTTCT 3730
Db 2756 CAAAAAGAAAAAAGCCCAATCAGTGTCTCTGACCTTGTCTTGTGATCCCTCAGTTTCT 2815
Qy 3731 TCTTGATTTCAACATGTGTGTGGGTTCTTAATTTGGGATAGATTAGCAAAATTTAACAT 3790
Db 2816 TCTTGATTTCAACATGTGTGTGGGTTCTTAATTTGGGATAGATTAGCAAAATTTAACAT 2875
Qy 3791 TGTGTTTGTGCTTACCCAGGGAATCCCACTTTCTGACTTGAAAGTACATGAAGAAGAA 3850
Db 2876 TGTGTTTGTGCTTACCCAGGGAATCCCACTTTCTGACTTGAAAGTACATGAAGAAGAA 2935
Qy 3851 TCCACAGAGTGTCTATCTGGCCAGATTTAAGTATTTCTATTTCTCTGTTCTCCCTCTCC 3910
Db 2936 TCCACAGAGTGTCTATCTGGCCAGATTTAAGTATTTCTATTTCTCTGTTCTCCCTCTCC 2995
Qy 3911 CTGAGAGCTCTTAATTTTAATTTGTCCCTCTCTTCTAGGTTAATTTCTCTTGAATTTGA 3970
Db 2996 CTGAGAGCTCTTAATTTTAATTTGTCCCTCTCTTCTAGGTTAATTTCTCTTGAATTTGA 3055
Qy 3971 GTTGAGAAAGAGTGTGACAGATATAGCAAAAGTTCCCAAGTGCACAAATTAAGTGTGT 4030
Db 3056 GTTGAGAAAGAGTGTGACAGATATAGCAAAAGTTCCCAAGTGCACAAATTAAGTGTGT 3115
Qy 4031 AAGAGTGTGGGGGAAAAATTAGCTTAATTTTCCCTACATGGAATCAACACTGTGAATTC 4090
Db 3116 AAGAGTGTGGGGGAAAAATTAGCTTAATTTTCCCTACATGGAATCAACACTGTGAATTC 3175
Qy 4091 AATCTTCAACTGAAGCCCTGCAGTCTCTCTAAACAATAGTGTGTGTTTCTTTTACA 4150
Db 3176 AATCTTCAACTGAAGCCCTGCAGTCTCTCTAAACAATAGTGTGTGTTTCTTTTACA 3235
Qy 4151 AAGTTTAAGCTATGTGTATTAATTAATAAAAAATTGCTGTGTCTACTTCAAGCTTTGT 4210
Db 3236 AAGTTTAAGCTATGTGTATTAATTAATAAAAAATTGCTGTGTCTACTTCAAGCTTTGT 3295
Qy 4211 TTTATGCCCATTTCAATATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4270
Db 3296 TTTATGCCCATTTCAATATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3355
Qy 4271 ATGTGTAAAAATGTTGTGTCTGTGTAATAATCTTAATAAGATCAATGTGTAATAATCTAT 4330
Db 3356 ATGTGTAAAAATGTTGTGTCTGTGTAATAATCTTAATAAGATCAATGTGTAATAATCTAT 3415
Qy 4331 TGTGCTGTT 4340
Db 3416 TGTGCTGTT 3425

RESULT 3
HSM807332 2745 bp mRNA linear PRI 30-AUG-2003
LOCUS HSM807332
DEFINITION Homo sapiens mRNA; cDNA DKF2p686P24239 (from c10ne DKF2p686P24239).
ACCESSION BX647188
VERSION BX647188.1 GI:34366216
KEYWORDS .

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2745)
AUTHORS Wamburt, R., Heubner, D., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
Fobo, G., Han, W., and Wiemann, S.
CONSRM The German Human cDNA Consortium
TITLE Direct Submision
JOURNAL Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuberberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOMA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZ66624239) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.
Location/Qualifiers
1..2745
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZ66624239"
/issue_type="human small intestine"
/clone_lib="686 (synonym: hlcc3). Vector psportl_sfi; host
DH10B; sites SfiIa + SfiIb"
/dev_stage="adult"
2707..2712
polYA_signal
polYA_site 2726

Query Match 62.7%; Score 2741.8; DB 9; Length 2745;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2743; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1615 ATGAGAGAGAGAGAGAGAGATGGATATTCAGAGATCAGAGAGAGAGAGAGAGATTTTACAGC 1674
DB 1 ATGAGAGAGAGAGAGAGAGATGGATATTCAGAGATCAGAGAGAGAGAGAGAGATTTTACAGC 60
QY 1675 GACTGACGCTCAGATATGATATTAAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1734
DB 61 GACTGACGCTCAGATATGATATTAAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 120
QY 1735 CATTTTCCCTGAGCCAGATGATGTAAGAGATTTGATGATTTACCCCTTCTGCTGTTG 1794
DB 121 CATTTTCCCTGAGCCAGATGATGTAAGAGATTTGATGATTTACCCCTTCTGCTGTTG 180
QY 1795 TGGCATTGGAGAGACATTACCAAAATTTATCTCCAGAGATTTTGGAGTACCTCGGTGG 1854
DB 181 TGGCATTGGAGAGACATTACCAAAATTTATCTCCAGAGATTTTGGAGTACCTCGGTGG 240
QY 1855 ATGAGAGAGAGAGAGATTCAGAGATTCAGAGAGAGAGAGAGAGATTCAGAGATTCAGC 1914
DB 241 ATGAGAGAGAGAGATTCAGAGATTCAGAGAGAGAGAGAGAGATTCAGAGATTCAGC 300
QY 1915 GGAATATGCTGTCTGAGCAAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1974
DB 301 GGAATATGCTGTCTGAGCAAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 360
QY 1975 AGTGGAGAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2034
DB 361 AGTGGAGAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 420
QY 2035 ACCCTTCACTCTCTGAGCTTTGTTTTCATTTACTGATTCAGAGATTCAGAGATTCAGC 2094
DB 421 ACCCTTCACTCTCTGAGCTTTGTTTTCATTTACTGATTCAGAGATTCAGAGATTCAGC 480
QY 2095 TCGGCTAATTTGATGATTTAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2154
DB 481 TCGGCTAATTTGATGATTTAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 540

QY 2155 GCTCTCAAGCTAAGATATAGATGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2214
DB 541 GCTCTCAAGCTAAGATATAGATGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 600
QY 2215 ATTTGCTTATGTTAAAGGGGTAAAGGGCTCTTCTTATGATGATGAGATGAGAGAGC 2274
DB 601 ATTTGCTTATGTTAAAGGGGTAAAGGGCTCTTCTTATGATGATGAGATGAGAGAGC 660
QY 2275 GCGCTTCCCTTAAAGAGCTGAGCTGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2334
DB 661 GCGCTTCCCTTAAAGAGCTGAGCTGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 720
QY 2335 TAGTGGATATGATGATTTTAAACCTTAAATTAAGAGAGAGATTCAGAGATTCAGAGATTCAGC 2394
DB 721 TAGTGGATATGATGATTTTAAACCTTAAATTAAGAGAGAGATTCAGAGATTCAGAGATTCAGC 780
QY 2395 GACCAAGAGAGAGATGACAGAGAGATTCAGAGAGAGAGAGAGAGATTCAGAGATTCAGC 2454
DB 781 GACCAAGAGAGAGATGACAGAGAGATTCAGAGAGAGAGAGAGAGATTCAGAGATTCAGC 840
QY 2455 ACATGGAGAGATGATGATTAACCTGCTGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2514
DB 841 ACATGGAGAGATGATGATTAACCTGCTGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 900
QY 2515 TGATACAGTATGATGAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2574
DB 901 TGATACAGTATGATGAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 960
QY 2575 AATTTCTTATGCTGCTCAAAACCAAAATCTTTGAGAGAGAGATTCAGAGATTCAGAGATTCAGC 2634
DB 961 AATTTCTTATGCTGCTCAAAACCAAAATCTTTGAGAGAGAGATTCAGAGATTCAGAGATTCAGC 1020
QY 2635 AGAAGAGATTAAGCTGATTAAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2694
DB 1021 AGAAGAGATTAAGCTGATTAAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1080
QY 2695 TATTAATGCTGATTAAGCTGATTAAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2754
DB 1081 TATTAATGCTGATTAAGCTGATTAAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1140
QY 2755 AGCAGAGATTCAGAGATTCAGAGAGAGATTCAGAGAGAGATTCAGAGATTCAGAGATTCAGC 2814
DB 1141 AGCAGAGATTCAGAGATTCAGAGAGAGATTCAGAGAGAGATTCAGAGATTCAGAGATTCAGC 1200
QY 2815 GGTGCTTCTGCTGCTGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2874
DB 1201 GGTGCTTCTGCTGCTGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1260
QY 2875 TTCTCATTCACTTGTAAATTAATTTGATGATGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2934
DB 1261 TTCTCATTCACTTGTAAATTAATTTGATGATGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1320
QY 2935 GTTTTGTAAATATATAGATTAAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 2994
DB 1321 GTTTTGTAAATATATAGATTAAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1380
QY 2995 ATTAATCTATGAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 3054
DB 1381 ATTAATCTATGAGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1440
QY 3055 CTTTCTTATGCTGCTGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 3114
DB 1441 CTTTCTTATGCTGCTGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1500
QY 3115 TAGGGCTTTTATCTTAAAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 3174
DB 1501 TAGGGCTTTTATCTTAAAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1560
QY 3175 GCCAAGATATGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 3234
DB 1561 GCCAAGATATGAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGC 1620

Hagob, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Mathew, C., McCarthy, M., Meldrum, J., Menus, L., Mihov, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Punnkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M. Direct Submission
Submitted (17-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 17, 2003 this sequence version replaced gi:27734042.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L10045
Center clone name: 749_I_16

FEATURES
source

Location/Qualifiers
1. 194780
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-749I16"
/clone_lib="RPC1-11 Human Male BAC"
2. 285
/rpt_family="L1MC3"
302..324
/rpt_family="(TTG)n"
529..637
/rpt_family="FLAM_A"
548..553
/note="<30 qual SINGL region"
complement(711..1007)
/rpt_family="AluSc"
complement(1038..1179)
/rpt_family="MER115"
2169..2473
/rpt_family="AluSp"
3642..3946
/rpt_family="AluY"
complement(4181..4391)
/rpt_family="AluJo"
4863..4901
/rpt_family="(GGGA)n"
5391..5716
/rpt_family="AluSc"
6700..6739
/rpt_family="(CA)n"
complement(7332..7491)
/rpt_family="MIR3"
7684..7805
/rpt_family="(CGGG)n"
7700..7856
/note="single clone coverage"
7788..7802
/note="<30 qual single clone coverage"
7825..7845
/note="<30 qual single clone coverage"
7931..7951
/rpt_family="GC_rich"
/rpt_family="GC_rich"

repeat_region	7980..8009	/rpt_family="GC_rich"
repeat_region	8388..8442	/rpt_family="GC_rich"
repeat_region	9670..9780	/rpt_family="L2"
repeat_region	10423..10557	/rpt_family="CT-rich"
repeat_region	11371..11675	/rpt_family="AluSc"
repeat_region	11724..11913	/rpt_family="MIR"
repeat_region	12708..13030	/rpt_family="AluJb"
repeat_region	13109..13189	/rpt_family="MIR3"
repeat_region	14044..14165	/rpt_family="CT-rich"
repeat_region	14184..14282	/rpt_family="CT-rich"
repeat_region	15860..15976	/rpt_family="CT-rich"
repeat_region	15894..16062	/rpt_family="CT-rich"
repeat_region	17522..17592	/rpt_family="C-rich"
repeat_region	18316..18420	/rpt_family="(CAGA)n"
repeat_region	21106..21188	/rpt_family="MIR"
repeat_region	22950..22950	/rpt_family="MIR"
repeat_region	23005..23172	/rpt_family="AluSc"
repeat_region	23173..23481	/rpt_family="MIR"
repeat_region	23482..23525	/rpt_family="AluJb"
repeat_region	23629..23697	/rpt_family="MIR"
repeat_region	24001..24113	/rpt_family="MIR"
repeat_region	24483..24483	/rpt_family="MIR"
repeat_region	24534..24534	/rpt_family="AluSc"
repeat_region	24569..24569	/rpt_family="AluSc"
repeat_region	24586..24632	/rpt_family="AluSc"
repeat_region	24770..24770	/rpt_family="MIR"
repeat_region	24821..25072	/rpt_family="AluSc"

Query Match 57.4%; Score 2511; DB 9; Length 194780;
Best Local Similarity 100.0%; Pred.No.0; Mismatches 0; Indels 0; Gaps 0;
Matches 2511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1832	GAATTTGAGCTACCTGTTGATGATGCGTACCGATGCGATTTGAGATCCAGAGAA	1891
Db	79250	GAATTTGAGCTACCTGTTGATGATGCGTACCGATGCGATTTGAGATCCAGAGAA	79309
Qy	1892	GCAACACCTCACCGGACGTTGGAATAGCTGTCTGCGACGAACCCCTGTCTTCAAT	1951
Db	79310	GCAACACCTCACCGGACGTTGGAATAGCTGTCTGCGACGAACCCCTGTCTTCAAT	79369

Db 81530 CAGTTCCTCTAAACATAGTGTGTTTCTTTTCTTAACAAGATTGAAGCTAGTGTAAATA 81589

Qy 41172 AATTAAAAAAATGCTGTCTGTCTACTTCACTTCACTTGTGTTTAAAGCCATTTTCATATG 4231

Db 81590 AATTAAAAAAATGCTGTCTGTCTACTTCACTTGTGTTTAAAGCCATTTTCATATG 81649

Qy 4232 TGTCTGTGTGTAATTCATTAATCTTTGATTCATTTCTGTGATGATGTAATTTGTTCTT 4291

Db 81650 TGTCTGTGTGTAATTCATTAATCTTTGATTCATTTCTGTGATGATGTAATTTGTTCTT 81709

Qy 4292 GTTAAATATCTTTAAAGAGTTCAATGTAATTAATTAATTTGTCGTTAA 4342

Db 81710 GTTAAATATCTTTAAAGAGTTCAATGTAATTAATTAATTTGTCGTTAA 81760

RESULT 5
AK055378 2296 bp mRNA linear PRI 30-JAN-2004

LOCUS AK055378 Homo sapiens CDNA FLJ30816 fls, clone FEBRA2001571.

ACCESSION AK055378

VERSION AK055378.1 GI:16550091

KEYWORDS oligo capping; fls (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakematsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuda, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirokawa, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Hosoda, M., Houta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takewaki, K., Arita, M., Imose, N., Musashino, K., Yuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshioka, Y., Matsunawa, H., Ichihara, T., Shishibata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, A., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsunaga, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagae, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs Nat. Genet. 36 (1), 40-45 (2004) 14702039

TITLE JOURNAL PUBMED REFERENCE AUTHORS

TITLE JOURNAL REFERENCE

unpublished

3 (bases 1 to 2296)

AUTHORS Isogai, T., Otsuki, T. and Sugiyama, T.

TITLE Direct Submission

JOURNAL Submitted (24-Oct-2001) Takao Isogai, Helix Research Institute, Genomic Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - 3' end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

source

1. .2296

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="FEBRA2001571"

/issue_type="brain"

/clone_lib="FEBRA2"

/dev_stage="fetus"

/note="Cloning vector: pME18SFL3"

ORIGIN

Query Match 52.5%; Score 2294.4; DB 9; Length 2296; Best Local Similarity 100.0%; Pred. No. 0; Matches 2295; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2024 TCGCTGATTAATACCTCTTCACTTCTCTGACTTGTGTTTCATTAATCTGATTTACAAA 2083

1 TCGCTGATTAATACCTCTTCACTTCTCTGACTTGTGTTTCATTAATCTGATTTACAAA 60

Qy 2084 AACTCTTCAATTCGGCTTAATGAGTTATGAGGCTGATTTGGATTTCTTTCCCTTT 2143

61 AACTCTTCAATTCGGCTTAATGAGTTATGAGGCTGATTTGGATTTCTTTCCCTTT 120

Qy 2144 TTGGAAATGGGCTCAAGCTAAAGCTATAGATGAGCAGATTCAGAACTTCAAGGGTC 2203

121 TTGGAAATGGGCTCAAGCTAAAGCTATAGATGAGCAGATTCAGAACTTCAAGGGTC 180

Qy 2204 TGTTCATTAATTCCTTATGTTAAAGGGGTAAGGGCTCTCTTCATTAATGACATG 2263

181 TGTTCATTAATTCCTTATGTTAAAGGGGTAAGGGCTCTCTTCATTAATGACATG 240

Qy 2264 AATATGAGAGCCCTTCTCTTAAAGCTGCTGCTGATGAGCCTCTTCAACCTGTA 2323

241 AATATGAGAGCCCTTCTCTTAAAGCTGCTGCTGATGAGCCTCTTCAACCTGTA 300

Qy 2324 CACCTCTTATATGAGGATAGTATTTTAAACCTTAAATTAACCAACCACTGACC 2383

301 CACCTCTTATATGAGGATAGTATTTTAAACCTTAAATTAACCAACCACTGACC 360

Qy 2384 ATGAGCTTATAGCAAGAGAGATGACAAAGTGAAGCAGTGAAGCAGCATTCTGACA 2443

361 ATGAGCTTATAGCAAGAGAGATGACAAAGTGAAGCAGTGAAGCAGCATTCTGACA 420

Qy 2444 GAGTGAAGAGACATGAGAGAGTTGGTATGATTAATCTGTGAAGAGATGTTGATTG 2503

421 GAGTGAAGAGACATGAGAGAGTTGGTATGATTAATCTGTGAAGAGATGTTGATTG 480

Qy 2504 AAATCTATTCGTATACAGTCATGAGGAGAGGATGTTGGCTGTGATTAATTTTTCAGT 2563

481 AAATCTATTCGTATACAGTCATGAGGAGAGGATGTTGGCTGTGATTAATTTTTCAGT 540

Qy 2564 TAAATGATTAATTTCTTACTGCTCAAAACCAAAATCTTTGAAAGAGAGAGTGGGA 2623

541 TAAATGATTAATTTCTTACTGCTCAAAACCAAAATCTTTGAAAGAGAGTGGGA 600

Qy 2624 TGGTATGTTTCAAGCAAGTATGAGCTGTTAAACAAAGACCTTAATTTGGAGTGGCAT 2683

601 TGGTATGTTTCAAGCAAGTATGAGCTGTTAAACAAAGACCTTAATTTGGAGTGGCAT 660

QY 2684 GCGAAAACCTGTAAATNGCTCTGTATGACATCACTTCTCAAGTATTCCTTCATTGGGC 2743
Db 661 GCGAAAACCTGTAAATNGCTCTGTATGACATCACTTCTCAAGTATTCCTTCATTGGGC 720
QY 2744 TTCATCTTTTGAAGAACTCTTGCGTGAGATAGACATTAAGGAGGATAGGGGAGA 2803
Db 721 TTCATCTTTTGAAGAACTCTTGCGTGAGATAGACATTAAGGAGGATAGGGGAGA 780
QY 2804 GTGTGAAATAGGTGCTTCTTGCGTGAGATAGACATTAAGGAGGATAGGGGAGA 2863
Db 781 GTGTGAAATAGGTGCTTCTTGCGTGAGATAGACATTAAGGAGGATAGGGGAGA 840
QY 2864 ACCTAATGAGCTTCCCATCACTTGTGAAAATTAATTTGTATGTGATCACTTGTGTCC 2923
Db 841 ACCTAATGAGCTTCCCATCACTTGTGAAAATTAATTTGTATGTGATCACTTGTGTCC 900
QY 2924 TCTCCCTCCCGTTTGTGAAAATATCAGGATAGCACTCCAGGACCTTGGTCTCAGT 2983
Db 901 TCTCCCTCCCGTTTGTGAAAATATCAGGATAGCACTCCAGGACCTTGGTCTCAGT 960
QY 2984 GTAGATCCCTATTAATCTATCTGAAAAGAAAATAGAGCAAGACCTGTGCTCAATAT 3043
Db 961 GTAGATCCCTATTAATCTATCTGAAAAGAAAATAGAGCAAGACCTGTGCTCAATAT 1020
QY 3044 ATAGAAATGCTTCTTCTTATGCTTCAAGACATTAATGTGTGAAAACAAAGTATGCTAT 3103
Db 1021 ATAGAAATGCTTCTTCTTATGCTTCAAGACATTAATGTGTGAAAACAAAGTATGCTAT 1080
QY 3104 CTCTAGAAAGTATGAGGCTTTTATCTTAAAGAAATATGTCCCGAGATTAATAGACTT 3163
Db 1081 CTCTAGAAAGTATGAGGCTTTTATCTTAAAGAAATATGTCCCGAGATTAATAGACTT 1140
QY 3164 TTAGAGAGAGACCAAGGATATGAGGCTGTGCGCCATCACTGAGACGAAAG 3223
Db 1141 TTAGAGAGAGACCAAGGATATGAGGCTGTGCGCCATCACTGAGACGAAAG 1200
QY 3224 AATAAGGATACCATTTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3283
Db 1201 AATAAGGATACCATTTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 3284 GTTTTGTGAGATGTGATCTGTGCTTCTGATTTGAATTTAAAGAAATATATCTGTG 3343
Db 1261 GTTTTGTGAGATGTGATCTGTGCTTCTGATTTGAATTTAAAGAAATATATCTGTG 1320
QY 3344 CAGCAGATGATATCTTGTGATGATCTTGTGCTTAAATTTCTCTTGTGTGTGTGT 3403
Db 1321 CAGCAGATGATATCTTGTGATGATCTTGTGCTTAAATTTCTCTTGTGTGTGTGT 1380
QY 3404 GT 3463
Db 1381 GT 1440
QY 3464 TCTATTAAGGAGACCTGTGAAAATCATTTGAGCAGAGATGTGAGAGAAATAGACTT 3523
Db 1441 TCTATTAAGGAGACCTGTGAAAATCATTTGAGCAGAGATGTGAGAGAAATAGACTT 1500
QY 3524 AATTCACACTAGGAGCTCTCATCTCACACCTTAAGAGAGAGATTTCTAGAAAACTGGCC 3583
Db 1501 AATTCACACTAGGAGCTCTCATCTCACACCTTAAGAGAGAGATTTCTAGAAAACTGGCC 1560
QY 3584 AGATTTTCTTGT 3643
Db 1561 AGATTTTCTTGT 1620
QY 3644 CTATGATATTTCTTCTGTAAGGTGCAAAAAAGAAAAAGCAATCAATGATCTCTTG 3703
Db 1621 CTATGATATTTCTTCTGTAAGGTGCAAAAAAGAAAAAGCAATCAATGATCTCTTG 1680
QY 3704 ACTTTGTCTTGTATCCCTCAGTTTCTTGTGATTTCAAGATGTGTGTGTGTGTGTGT 3763
Db 1681 ACTTTGTCTTGTATCCCTCAGTTTCTTGTGATTTCAAGATGTGTGTGTGTGTGTGT 1740
QY 3764 TGGGATATGATTAAGCAATTTAACTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3823

Db 1741 TGGGATATGATTAAGCAATTTAACTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1800
QY 3824 TTCTGACTTGAAGTACGTGAGAGAGATTCACAGAGTGTCTATCTGGCAGATTTAAGT 3883
Db 1801 TTCTGACTTGAAGTACGTGAGAGATTCACAGAGTGTCTATCTGGCAGATTTAAGT 1860
QY 3884 ATTCTATTTCTTGT 3943
Db 1861 ATTCTATTTCTTGT 1920
QY 3944 AGGTTAATTCCTTGTGATTTGATCTTGTGTGAAGAGGTTGACAGTATGACAA 4003
Db 1921 AGGTTAATTCCTTGTGATTTGATCTTGTGTGAAGAGGTTGACAGTATGACAA 1980
QY 4004 GTTCCAGAGCAAAATTAACAGTGTGTGAGAGTGTGGGAGAAAATTAATGATTTTTC 4063
Db 1981 GTTCCAGAGCAAAATTAACAGTGTGTGAGAGTGTGGGAGAAAATTAATGATTTTTC 2040
QY 4064 CTACATGGGATACCACTGTGAAATTCATCTTCAACTGAGAGCCCTGCAATTTCTTAA 4123
Db 2041 CTACATGGGATACCACTGTGAAATTCATCTTCAACTGAGAGCCCTGCAATTTCTTAA 2100
QY 4124 AACATAGT 4183
Db 2101 AACATAGT 2160
QY 4184 TTGCTGT 4243
Db 2161 TTGCTGT 2220
QY 4244 AATTCATTAATTTGTGATACCATTTCTGATGTGTAAATGTGTGTGTGTGTGTGT 4303
Db 2221 AATTCATTAATTTGTGATACCATTTCTGATGTGTAAATGTGTGTGTGTGTGTGT 2280
QY 4304 TAAAGATTCATTTGT 4319
Db 2281 TAAAGATTCATTTGT 2296

RESULT 6
AK124185
LOCUS 2638 bp mRNA linear PRI 30-JUN-2004
DEFINITION Homo sapiens cDNA FLJ42191 f18, clone THYM0203053.
ACCESSION AK124185
VERSION AK124185.1 GI:34529914
KEYWORDS oligo cloning; f18 (Full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Oobayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, U., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nishigaki, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, K., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuma, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirooka, S., Chiba, Y., Iehida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Houchi, T., Kusano, U., Kanehori, K., Takahashi, F., Hara, H., Tanabe, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,


```

Db      1979 AAAATAGACCAAGACCTCTGCTCAATATATAGAAATGCGCTTCTTAGTCTTCAG 2038
Qy      3072 GACTATTGTGTAAAAAAGTAGGGGTCTAATCTCTAAGAGTAGGGCTTTATCCCT 3131
Db      2039 GACTATTGTGTAAAAAAGTAGGGGTCTAATCTCTAAGAGTAGGGCTTTATCCCT 2098
Qy      3132 AAGAGAAATATGTCCCAAGATTATAGCACTTTTATAGAGAAAGCAAGGTATGTGGGT 3191
Db      2099 AAGAGAAATATGTCCCAAGATTATAGCACTTTTATAGAGAAAGCAAGGTATGTGGGT 2158
Qy      3192 GTGTGGCTGGCCATCATGATGAGACGAAAGAGAAATGGATACATTTGTGGAGAGAA 3251
Db      2159 GTGTGGCTGGCCATCATGATGAGACGAAAGAGAAATGGATACATTTGTGGAGAGAA 2218
Qy      3252 GAAAAGTTCTCTAGAGGGGCTCCCACTGCTTAAAGTTTGTGAGAGTTGATCTTGCTT 3311
Db      2219 GAAAAGTTCTCTAGAGGGGCTCCCACTGCTTAAAGTTTGTGAGAGTTGATCTTGCTT 2278
Qy      3312 CCTGATTTGACTTTTAAAGAAATTTATCTGGCAGACATGTAGTATCTTGATGATCT 3371
Db      2279 CCTGATTTGACTTTTAAAGAAATTTATCTGGCAGACATGTAGTATCTTGATGATCT 2338
Qy      3372 TGCTGCTTTATTTCTCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3431
Db      2339 TGCTGCTTTATTTCTCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2398
Qy      3432 TTGTAACTCACTCTGCTTTAGAGAGAGGGCTCTCTAATAAGGAACTGCTGTAACTTCA 3491
Db      2399 TTGTAACTCACTCTGCTTTAGAGAGAGGGCTCTCTAATAAGGAACTGCTGTAACTTCA 2458
Qy      3492 TTGCAGCAGAGATGTAGAGAGAAATAGGACTTAATTCACCTAGGGCTCTCATCTCACAC 3551
Db      2459 TTGCAGCAGAGATGTAGAGAGAAATAGGACTTAATTCACCTAGGGCTCTCATCTCACAC 2518
Qy      3552 CTTAAGAGAGAGATTTCTAGAAAAAAGCTGGGCCAGATTTCTTTGCTTCATCATTTTAA 3611
Db      2519 CTTAAGAGAGAGATTTCTAGAAAAAAGCTGGGCCAGATTTCTTTGCTTCATCATTTTAA 2578
Qy      3612 TGTGGAGAGCTGTTCAGTTTCTTACTCTTACTTATGTGATTTCTTGTAAAGTGTCC 3671
Db      2579 TGTGGAGAGCTGTTCAGTTTCTTACTCTTACTTATGTGATTTCTTGTAAAGTGTCC 2638

RESULT 7
BC058629 2933 bp mRNA linear ROD 30-JUN-2004
LOCUS Mus musculus RIKEN cDNA 4121402D02 gene, mRNA (cDNA clone MGC:64809
DEFINITION IMAGE:6402215), complete cds.
ACCESSION BC058629
VERSION BC058629.1 GI:35193129
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2933)
Strausberg,R.D., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,J., Shennan,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buettow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,L., Martusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaeetz,T.E., Brownstein,M.J., Uedlin,T.B., Toshiyuki,S.,
Carinici,P., Prange,C., Raha,S.S., Loqueano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Morley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallus,D.E.,

```

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2933)
Strausberg,R.
Direct Submission
Submitted (22-SEP-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas L. Casavant.
Web site: <http://genome.uiowa.edu>
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
Schaeetz,T., Smith,C., Smitr,E., Tack,D., Trout,K., Walters,J.,
Casavant,T., Soares,M.B.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: <http://image.liml.gov>
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21311987.
Location/Qualifiers
1..2933
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:64809 IMAGE:6402215"
/tissue_type="mouse, brain 12.5 dpc"
/clone_1ib="NIH BMAP_FCO"
/lab_host="DH10B"
/note="Vector: pYX-Asc"
1..2933
/gene="4121402D02R1k"
/note="synonym: 2810017F12R1k"
/db_xref="GeneID:74026"
/db_xref="MGI:1921276"
63..1913
/gene="4121402D02R1k"
/codon_start=1
/product="RIKEN cDNA 4121402D02"
/protein_id="AAH58629.1"
/db_xref="GI:35193130"
/db_xref="LocustID:74026"
/db_xref="MGI:1921276"
/translation="MTMSAVFKAAAPAGNPEORLDYERAAALGPEDESGAAEAH
FLPRHKLKEPPLASGSGSPSPAGCGGKGRGLIPGAAPGQDEBWSGVP
LPCPPATKQAGIGEYVAGSPAGSPRYQAVLPITGSIYVAAKEPTPWAGDGG
AAPPAATKADPAGPPPLPGPPPLATAGTLAASBRWKSIRKSPLAGGGSGAS
SOAALCKQLILDLLEQQOOLQAKKEKEISERPTLLARIEMRMQVARD
NKEHKLQGYEPERESAEISSEKIKARQPELCEISQALSPKSPSCSGSGKHK
TFEGTERKTVPVKAPERSKVXTTPKHSPIKEPPCGSISTVCKRELRSQETPEK
RSVDVTPPLSTPQKSPSTHPKPKAPSSMBDLPYSTTEMILCRHKEPTPLPARE
SSPKBEETVAPKSSVAGESVLAPEMRDHSVLEPLADPSPDILEXLDVSFSKR
HAKLIDERRRRKRIORIRREORITIORLQRYKKKGIOESRPVTPPEPDVDSL
LITPLPVAVAFRPLPKAPQNFELPMDERSKLEIQQKTHPRTRCK"

ORIGIN
Query Match 44.6%; Score 1948.2; DB 10; Length 2933;
Best local Similarity 83.9%; Pred. No. 0;
Matches 2463; Conservative 0; Mismatches 378; Indels 98; Gaps 21;

QY	10	GGACACTGACCCCTTCCACACCCCTCTCCGCGCTCGGTGCGGAGCGCTGTCCGACACA	75
Db	1	CGACCTGCCCCCTTCCACCCCTCTCTCGTCTCGTCTCCCGCGCTGTCCGACACA	60
QY	76	CTATGACCATGATGCCGCGTGTTCAAAGCGGCGCGGACCCCTCGCGCGGCAATCTTG	135
Db	61	CTATGACCATGATGCCGCACTGTTCAAAGCGGCGCGGACCCCTCGCGCGGCAATCTTG	120
QY	136	AGCAGCGACTGAGCTACGAGCGGCGCTGCGGCGCTGCGGCGCGGACGAGCTTGAGG	195
Db	121	AGCAGCGACTGAGCTACGAGCGGCGCTGCGGCGCTGCGGCGGCGGACGAGCTTGAGG	180
QY	196	CGGCGGAAGCCCACTTCTCTCCCGCGGACCGTAAAGCTCAAGAACCGGGGCCCCCTGG	255
Db	181	CGGCGGAAGCCCACTTCTCTCCCGCGGACCGTAAAGCTCAAGAACCGGGGCCCCCTGG	240
QY	256	CCTCTCTCCAGAGGCGGAGCCCGGCGCTTCCCGGCGCGCT---GGGCGGACAGGAGCC	312
Db	241	CCTCTCTCCAGAGGCGGAGCCCGGCGCTTCCCGGCGCGCT---GGGCGGACAGGAGCC	300
QY	313	GGGCGTGTGTACTCCCGGCGGCGGCGCGCGCGCGCGCGGACAGAGAGAGCTGGGCGGTT	372
Db	301	GGGCGTGTGTACTCCAGCGGCGGCGGCGCGCGCGCGCGGACAGAGAGAGCTGGGCGGTT	360
QY	373	CGGTGCGCTTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	432
Db	361	CGGTGCGCTTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	420
QY	433	CGGACGCGGAGCGGCGCTGCAAGCCCGGCGCGCAAGTATCAGGCGGTGTGCTTCA	492
Db	421	TGCGACGCGGCGCTGCGCTGCAAGCCCGGCGCGCAAGTATCAGGCGGTGTGCTTCA	480
QY	493	CGGCGCTCTCTC---GTGGCGGCGGCGCAAGAGCTTACGCGCTTGCGGCTTGCGGACAAGG	549
Db	481	CGGCGCTCTCTCTCTG	540
QY	550	GGGCGGCGCTCCCCGCGCTGCAAGCGCTTGTGAGCCCGGCGGAGCCCCCACTACTTTC	609
Db	541	GGGCGGCGCTCCCCGCGCTGCAAGCGCTTGTGAGCCCGGCGGAGCCCCCACTACTTTC	600
QY	610	CGGCGCGGCGCAACCCCTGCGGCGCGCAACCGCGCGGACCTGTGCGGCGGCGGAGGCA	669
Db	601	CGGCGCGGCGCAACCCCTGCGGCGCGCAACCGCGCGGACCTGTGCGGCGGCGGAGGCA	660
QY	670	GATGGAAGATGATGAGGAAGAGCCCTCTCGGCGGCTGTGTGTGTGTGTGTGTGTGTG	729
Db	661	GATGGAAGATGATGAGGAAGAGCCCTCTCGGCGGCTGTGTGTGTGTGTGTGTGTGTG	720
QY	730	AGGCGCGCTGCTTCAACAGATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	789
Db	721	AGGCGCGCTGCTTCAACAGATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	780
QY	790	AGCAGCTGCAAGCCCAAGAAAAGAGATCGAGAGCTGAATCGAGAGAGACAGCTTC	849
Db	781	AGCAGCTGCAAGCCCAAGAAAAGAGATCGAGAGCTGAATCGAGAGAGATCGCTTC	840
QY	850	TTGCTCGGATTTGAACGTATGAAAAGCGGAGTACAGCTGTAAAGAGGATTAACAGAAAG	909
Db	841	TTGCTCGGATTTGAACGTATGAAAAGCGGAGTACAGCTGTGTAAAGAGGATTAACAGAAAG	900
QY	910	AAAGGCAACAGCTGTTTCAGGCTATGAAACTGAAGAGAGAGAGAAAACAGAGCTATCTG	969
Db	901	AAAGGCAACAGCTGTTTCAGGCTATGAACTGAAGAGAGAGAGAAAACAGAGCTATCTG	960
QY	970	AGAAAATTAACTGGAATGTGCAAGCGGAGCTTTTCGAGACATCTCCAGACTCTGCTTCCA	1029
Db	961	AGAAAATTAAATTTGAGAGCGGACAGCGGAGCTTTTGAGACATCTCCAGGCTCTGCTTCCA	1020
QY	1030	AGCCCTTCTCATGTGTGCGGAGTGTGAAAGGAGACATTAAGAGAAAATCCCATTTTGAAGTA	1089
Db	1021	AGCCCTTCTCATGTGTGCGGAGTGTGAAAGGAGACATTAAGAGAAAATCCCATTTTGAAGTA	1080
QY	1090	CAGAAAGAAAGACTCTGTATAAAAGCTGCTCTGAATTTTCAAAAGTCAAAACAAA	1149

Db	1081	CAGAAAGAAAGACTCCCTGTTAAAAAGCTGGCTCTGAATTTTCAAAAGTCAAAAACAAAA	1140
Qy	1150	CTCCTAAGCACTCTCTATTAAAAGAGAACCTGTGGTTCCTTATCTGAACCTGTTTGA	1209
Db	1141	CTCCTAAGCACTCTCCCATTTAAAGAGAACCTGTGGTTCATATACGAATCTGTTTGA	1200
Qy	1210	AACGTGAATTGAGAGCCAGAAACCACAGAAAACCCCGCTCTTCAGTGAACACCCAC	1269
Db	1201	AACCTGAATTGAGAGCCAGAAACCACAAAAAGCCCCGCTCTTCAGTGAATACCCAC	1260
Qy	1270	CAAGACTCTCACTCCCCCAAAAGGAGCCAGAACCCATCCCAAGAGAAAGCCCTTCAA	1329
Db	1261	CAAACACTCTGACTCCCCCAAAAGGAGCCAGAACCCACCCCAAGAGAAAGCCCTTCAA	1320
Qy	1330	GTCGATAGAGAAAGATTGGCCGTACCTTTCCACACAGAAATGTATTTGTGTGTGGACCC	1389
Db	1321	GTGATAGAGAAAGATTGGCCGTACCTTTCCACACAGAAATGTATTTGTGTGTGGACCC	1380
Qy	1390	AGCCTCCCCCATCACTGTTACATTAAGGGAATCTCTCCAAAGAGAGAGACTGTAG	1449
Db	1381	AGCCTCCCCCATCACTGTATCAATTAAGGGAATCTCTCCAAAGAGAGAGACTGTAG	1440
Qy	1450	CAAGGTCTGTAGTCCATCAAGTGTTCAGAGAAACTTCAGCTTGGCTGTCTCTT	1509
Db	1441	CAAGGTCTGTAGTCCATCAAGTGTTCAGAGAAACTTCAGCTTGGCTGTCTCTT	1500
Qy	1510	GGAAGGACCACTCAGTAGAGCTCTTAAGGAGCCAAATCTTCAGACCTTTTGGAGAAC	1569
Db	1501	GGAAGGACCACTCAGTAGAGCTCTTAAGGAGCCAAATCTTCAGACATTTTGGAGAAC	1560
Qy	1570	TGATGACAGTGTGTGTTTGAAGGCGCATCAAAATCTGAGCTGTATGAGAGAGAA	1629
Db	1561	TGATGACAGTGTATTTTCAAGAGGCGATCAAAATCTGAGCTAGAGAGAGAGAA	1620
Qy	1630	AAAGATGGGATTTCAAGAGATCAGGGAACAAAGAAATTTTACACGCACTGACGTCGAA	1689
Db	1621	AAACATGGGATTTCAAGAGATCAGGGAACAAAGAAATTTTACACGCACTGACGTCGAA	1680
Qy	1690	TGTTAAAAAGAAAGAAATTCAGGAATCTGAGCTGAGGTTACCTCAATTTTCCCTGAGC	1749
Db	1681	TGTTAAAAAGAAAGAAATTCAGGAATCTGAGCTGAGGTTACCTCAATTTTCCCTGAGC	1740
Qy	1750	CAGATGATGTTGAAGTTGATGATTAACCCCTTCTTCCCTGTGTGATTTGAGAGAC	1809
Db	1741	CAGATGATGTTGAAGTTGATGATTAACCCCTTCTTCCCTGTGTGATTTGAGAGAC	1800
Qy	1810	CATTACCAAAATTAACCTCCACAGAAATTTGAGCTACCTGTGTGATGAGCCGAT	1869
Db	1801	CATTGCCAAAATTAACCTCCACAGAACTTTGAACCTACCTGTGTGATGAGCCGAT	1860
Qy	1870	GCAATTTGGAGATCCAGAAAGGAAACACTCAACGGAAGGTGAGGAAATAGCTGTGCT	1929
Db	1861	GCAGTTGGAGATCCAGAAAGGAAACACTCAACGGAAGGTGAGGAAATAGCTGTGCT	1920
Qy	1930	GGCAGAGAACCTGCTCTCAGATAGTTTATGACATGCCATTTCC-GGAGGTGGCAGAGACT	1988
Db	1921	GGC-GGTAGCTGTCTCAGATAGTTTATGACATGCTGTTCCGAGAGAAATGACAGAGACT	1979
Qy	1989	GTATATGTGAACCTTTGTCTCACATATGTTATCATCTGCTGATATATACCTTTATATCTT	2048
Db	1980	GTATGTGTGAACGTGTGCTCATGTCTGTATATCATCATGTATATCCTTTTCATCTCTC	2039
Qy	2049	CCTTGA-----CTTGTGTTCACTATCTGATTTTCAAA-----AAACTCTT	2090
Db	2040	CTTTATATTAATGTTTGTATTTCTTATCTGATTTTCAAAACAAACAAAGAAAGCTCT	2099
Qy	2091	TCATTCGGCTAATTTGTAGTTATGAGAGGTGATTTGGATTTCTTTTCCTTTTTGGGAA	2150
Db	2100	TGCGTTTGGCTGATTTGTGAATTTA---GGGGTTACTGAGAGGCGCTTTTCCATTTTGGGAG	2156
Qy	2151	ATGGGCTCTCAAGCTAAAGCTATGAGTGCAGATTCAGAACTTCAAGGGCTCTGTTTCT	2210

Db	2157	ATGAG--CTTAGCTCGCTCTAATAGG--GGAGAGACTTGGGGGTCCCAAGGGCCACTTGT	2212
Qy	2211	ATACATTTGCTTATGTTAAAGGGGTAAAGGGCTCTCTTCAATTAGACATGTGGAAGATGA	22707
Db	2213	ACTCACCTGCTGCTCTAACGGGGGTAAA-----GGGCTCTTCATTAG--CGTGTCTAAGATGG	2267
Qy	2271	AGCAGCCCCCTTCTTTAAGCTGTGCTTCATAGGACACTCTTCTACACCTTGGTACACCTC	23303
Db	2268	AGCAACCCCTTCTCTGAGAGCTGTGCTCCCTGACCCCGTGCCTCCCTGTGGCGCTTC	23277
Qy	2331	CTTATAGGGGTATGTGATTTTAACTTAACTAAATAAACAAACACTCACCATAGACT	23909
Db	2328	TGAATTAGTGTGGGTGTAGTAGTT-----TCACTGTAAGT	23635
Qy	2391	TTAGAGCCAGAAAGGAATGACAAAGTGAAGCGATGACCAAGCATCTTCAACAGATGAG	24500
Db	2364	TTAGGGGCCCGAGGAATA-----GCATGAAGTATGAAACCAAGACACTTCTCAGAGGGGA	2418
Qy	2451	AAAGACATCGAAGAGTGTGTAGATTAAGTCTGAAAAGATAGTGTTCATTTGAAACTAT	25101
Db	2419	AACGCGAAG-----TTTTACCTGAAGTTGTCTCTTTAAACTGC	24604
Qy	2511	TCTGATACAGTCAATGAGGAAGGATGTTGGCTGATTAATTTTTCAGTTAATGGA	25707
Db	2461	TCTGTGAACGTGGCATGTGGAGGAGACGTGCAGCTG--GTTCATCTTCTTAGTTGATGG	2519
Qy	2571	TAAC--AAATTCCTTACTGTCTCAAAAACAAATCTTTGAAAAAGAAAGTGGGATGCT	2627
Db	2520	TAACTGAAGTTCTCTTCCACACCCCAACAAAGAAAGATGTTTGAAGAAAGATTTGGAGTGT	25797
Qy	2628	TAGTTTCAGAAACAAGTTACAGCTGTAA--CAAAAGCACTTAGTATTTGGATGCGATGCC	2686
Db	2580	AAGTTTCCAACTAGTTAAACAGTGTAAACCGATAGAGCTTCATTTGGGGATGCGATGCC	26393
Qy	2687	AAAACCTGTAAATANGTCTGTATCAATCAATCACTTCAAGTATCTTCAATGGGCTTC	2746
Db	2640	AGAGTGTGTAGAGATGTCTG--TCATGTCACTTTCAAGGATTCCTTAGTGGGGCTTG	26977
Qy	2747	ATCCTTTAGAGAACTCTTGGTGTGGGATGAGACTTAGGAGAG--TAGGGGAGAG	2804
Db	2698	ATTCATATAGCCAAACTTTTGGTGTGAGTGAAGATTTGGGGTGGGCCCGGGGTAGGG	275
Qy	2805	TGTGAAATAGGTGCTTCTCTTTGGCTGGCAAAATGTCATCATTTGAAACAAACAGATGA	2864
Db	2758	TATGAAAGATATGCTTCTCTGCTGCGACAGTGTCTACAGCTTGAAACAAACAGTGTGA	2817
Qy	2865	CCTAATAGGCTTCTCCATTCATCTTGTAAAAATA-ATTGATGTGTACCATCTTGTGCC	2922
Db	2818	CCTAATAGAACTT--TATTCGTTTAGTGAATAATAGATTATATACGAACCACTTGTGTTTC	2874
Qy	2924	TCTCCCTCCCGTTTGT 2942	
Db	2875	CCATCTCTGCTGCTTGT 2893	
RESULT 8			
AR379622			
LOCUS			
DEFINITION Sequence 167 from patent US 6607879.			
VERSION AR379622			
KEYWORDS AR379622.1 GI:40087256			
SOURCE			
ORGANISM			
Unknown.			
REFERENCE			
1 (bases 1 to 1545)			
AUTHORS Cooks,B.G., Stuart,S.G. and Seilhamer,J.J.			
TITLE Compositions for the detection of blood cell and immunological			
response gene expression			
PATENT: US 6607879-A 167 19-AUG-2003;			
JOURNAL			
TEATRES			
Location/Qualifiers			
1. 1545			
/organism="unknown"			

ORIGIN	/mol_type="genomic DNA"
Query Match	34.5%; Score 1508.4; DB 6; Length 1545;
Best Local Similarity	97.7%; Pred. No. 3,56-278;
Matches 1509; Conservative	0; Mismatches 36; Indels 0; Gaps 0;
Db	2796 GGGGAGAGTGTGGAAATAGTGCTCTCTTGGCTGGCAATGTCTACATCTTGAAACAA 2855
1	GGGGAGAGTGTGGAAATAGTGCTCTCTTGGCTGGCAATGTCTACATCTTGAAACAA 60
Db	2856 AACAATGTAAGTATGAGTCTCTCACTTCCTTGTAATAATATTGTATGTATCAAT 2915
61	ACAGATGTAAGTATGAGTCTCTCACTTCCTTGTAATAATATTGTATGTATCAAT 120
2916	CTTGCTCTCTCCCTCCCGTCTTGTGTAATAATACAGATAGCACTCCAGGCACTTTG 2975
121	CTTGCTCTCTCCCTCCCGTCTTGTGTAATAATACAGATAGCACTCCAGGCACTTTG 180
2976	GTCTCAGTGAAGATCCCTATTTAACTATCTGAAAGAAAATAGAGCAAGACCTGTGTC 3035
181	GTCTCAGTGAAGATCCCTATTTAACTATCTGAAAGAAAATAGAGCAAGACCTGTGTC 240
3036	TCAATATATAGAAATGCTCTTTCTTTAGTCTTACAGACTATGTGTGAAACAAATAGG 3095
241	TCAAATATATAGAAATGCTCTTTCTTTAGTCTTACAGACTATGTGTGAAACAAATAGG 300
3096	GATCTAATCTCCCTAGAAAGTGGGGCTTTTATCTTAAAGAAATATGTCCCGACATTTAT 3155
301	GATCTAATCTCCCTAGAAAGTGGGGCTTTTATCTTAAAGAAATATGTCCCGACATTTAT 360
3156	TAGCACTTTTAGAGAGAGAGCCAAAGTATGAGGGTGTGTGCTGTGCGCCCATCAGTGAGC 3215
361	TAGCACTTTTAGAGAGAGAGCCAAAGTATGAGGGTGTGTGCTGTGCGCCCATCAGTGAGC 420
3216	ACGAAGAGAAATGGGATACCATTTGTGGAAAGAAAGAAAGTTCTTCAGGGGCTCCCA 3275
421	ACGAAGAGAAATGGGATACCATTTGTGGAAAGAAAGAAAGTTCTTCAGGGGCTCCCA 480
3276	CTGCTAAAGTTTTTTTGAGATGTGTGATCTGTGCTCCTGGAATTGACTTTTAAAGAAAT 3335
481	CTGCTAAAGTTTTTTTGAGATGTGTGATCTGTGCTCCTGGAATTGACTTTTAAAGAAAT 540
3336	TATTCGTGACACACATAGTATTCCTTGATGATCTGTGCTCTTATTTCTCCTTTTGT 3395
541	TATTCGTGACACACATAGTATTCCTTGATGATCTGTGCTCTTATTTCTCCTNNNN 600
3396	GT 3455
601	NN 660
3456	GTGGGCTCTATATAAGGAGACCTGTGTAACCTCATTTGACGAAAGATGTAGAGAGAA 3515
661	GTGGGCTCTATATAAGGAGACCTGTGTAACCTCATTTGACGAAAGATGTAGAGAGAA 720
3516	TAGACTTAATTTCACTAGAGGCTCTCATCTCACACCTTAAAGAGAGATTTCTAGAAA 3575
721	TAGACTTAATTTCACTAGAGGCTCTCATCTCACACCTTAAAGAGAGATTTCTAGAAA 780
3576	ACTGGGACAGATTTTCTTTGTTCATCATTTTAAATGTGGCAGGCTGTTCAGTTTTCTT 3635
781	ACTGGGACAGATTTTCTTTGTTCATCATTTTAAATGTGGCAGGCTGTTCAGTTTTCTT 840
3636	ACTCTTACCTATGTGATATTTCTTCGTATACCTGTCCAAAAAGAAAAAGAACCCATCAGT 3695
841	ACTCTTACCTATGTGATATTTCTTCGTATACCTGTCCAAAAAGAAAAAGAACCCATCAGT 900
3696	GTCTTTGACTTTGTTCTTTGATCCCTCAGTTTCTTCTGATTTACAGATGTGTGGGTT 3755
901	GTCTTTGACTTTGTTCTTTGATCCCTCAGTTTCTTCTGATTTACAGATGTGTGGGTT 960
3756	CTTAATTTGGGATGTAGTAAAGAAATTAACATGTGTGTGTGCTTACCCAGGGGAC 3815

Db	961	CTAATATTTGGGATGAGAGTTAGCAAAATTAAACAATGCTGTTTGAGCCCTACCACAGGGGAC	1020
QY	3816	TCGCCAGTTTCTGAAGTGAAGTGAAGAAATCCAGAGGTGCTATCTGGCCAGAT	3875
Db	1021	TCGCCAGTTTCTGAAGTGAAGTGAAGAAATCCAGAGGTGCTATCTGGCCAGAT	1080
QY	3876	TTAAGTAGATCTAATTCCTGGTGTCTCCCTCCCTGAGGACCTCTAATTTTAATGTC	3935
Db	1081	TTAAGTAGATCTAATTCCTGGTGTCTCCCTCCCTGAGGACCTCTAATTTTAATGTC	1140
QY	3936	CCTCTTCAGGTAAATCTCCCTTGATTTGAATTGTGTGAAGAAGAGGTGGAACAGTGA	3995
Db	1141	CCTCTTCAGGTAAATCTCCCTTGATTTGAATTGTGTGAAGAAGAGGTGGAACAGTGA	1200
QY	3996	TTAGCAAAAGTTCCAAAGTCAAAATTTACAGTGTGTGAAGTGTGGGGGAAAAATTAGCTT	4055
Db	1201	TTAGCAAAAGTTCCAAAGTCAAAATTTACAGTGTGTGAAGTGTGGGGGAAAAATTAGCTT	1260
QY	4056	ATTTTTCCTTCATAGGGAATACAAACCTGTGAATTCATCTTCAACTGAAGGCCCTGCAGT	4115
Db	1261	ATTTTTCCTTCATAGGGAATACAAACCTGTGAATTCATCTTCAACTGAAGGCCCTGCAGT	1320
QY	4116	TCTCCTAAAAATAGTGTGTTGTTTCTTTTAAACAAAGTTTAAAGCTAGTGTAAATTAAT	4175
Db	1321	TCTCCTAAAAATAGTGTGTTTCTTTTAAACAAAGTTTAAAGCTAGTGTAAATTAAT	1380
QY	4176	AAAAAAAATGCTGTCTGTCTACTTCAGCTTGTGTTTATGCCCAATTCATATTTGTGTC	4235
Db	1381	AAAAAAAATGCTGTCTGTCTACTTCAGCTTGTGTTTATGCCCAATTCATATTTGTGTC	1440
QY	4236	TGTGTGTGAATTCATTAATCTTTTGATACATTTCTGATGTGTAAATTTGGTGTCTTGTA	4295
Db	1441	TGTGTGTGAATTCATTAATCTTTTGATACATTTCTGATGTGTAAATTTGGTGTCTTGTA	1500
QY	4296	ATATCTTAATAAGAGTCAATTTGTAAATAAACTAATGTGGCGCTT	4340
Db	1501	ATATCTTAATAAGAGTCAATTTGTAAATAAACTAATGTGGCGCTT	1545

RESULT 9					
BD136391					
LOCUS	BD136391	1509 bp	DNA	linear	PAT 18-SEP-2002
DEFINITION	95 human secretory proteins.				
ACCESSION	BD136391				
VERSION	BD136391.1	GI:23231336			
KEYWORDS	JP 2002506627-A/78.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 1509)				
AUTHORS	Ruben,S.M., Ni,U., Rosen,C.A., Yu,G.L., Young,P.E., Fen,P., Soppet,D.W., Wei,Y.F., Endress,G.A., Duan,R.D., Kyaw,H., Ebner,R., Lalleur,D.R., Olsen,H.S., Shi,Y. and Moore,P.A.				
TITLE	95 human secretory proteins				
JOURNAL	Patent: JP 2002506627-A 78 05-MAR-2002;				
COMMENT	HUMAN GENOME SCIENCES INC				
	OS Homo sapiens (human)				
	PN JP 2002506627-A/78				
	PD 05-MAR-2002				
	PF 18-MAR-1999 JP 2000536733				
	PR 19-MAR-1998 US 60/078566,19-MAR-1998 US 60/078576 PR				
	19-MAR-1998 US 60/078573,19-MAR-1998 US 60/078574 PR				
	19-MAR-1998 US 60/078577,19-MAR-1998 US 60/078578 PR				
	19-MAR-1998 US 60/078581,19-MAR-1998 US 60/078577 PR				
	19-MAR-1998 US 60/078563,01-APR-1998 US 60/080314 PR				
	01-APR-1998 US 60/080312,01-APR-1998 US 60/080313 PI				
	M RUBEN,JUAN NI,CRAIG A ROSEN,GUO				
	LIANG YU,PAUL E YOUNG,				
	PI PING FENG,				
	PI DANIEL R SOPPET,YING FEI WEI,GREGORY A ENDRESS,ROXANNE D DUAN,				
	PI HUA KYAW,				
	PI REINHARD EBNER,DAVID W LALLEUR,HENRIK S OLSEN,YANGU SHI,PAUL				

```

PI      A MOORE
PC      C12N15/09, A61K38/00, A61K48/00, A61P43/00, C07K14/47, C07K16/18,
PC      C12N1/15,
PC      C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12Q1/68, G01N33/53, G01N33/
566,
PC      C12N15/00, A61K37/02, C12N5/00
CC      95 human secretory proteins
FH      Key
FT      Location/Qualifiers
FT      source          1..1509
                        /organism="Homo sapiens (human)".

FEATURES
Source
Location/Qualifiers
1..1509
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match          33.8%; Score 1478; DB 6; Length 1509;
Best Local Similarity 99.9%; Pred. No. 2.3e-272;
Matches 1500; Conservative 0; Mismatches 0; Indels 2; Gaps 2

```

FEATURES	source	Location/Qualifiers
		1..1509
		/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
ORIGIN		
Query Match	33.8%; Score 1478; DB 6; Length 1509;	
Best Local Similarity	99.9%; Pred. No.2.3e-272;	
Matches 1509; Conservaive	0; Mismatches 0; Indels 2; Gaps 2	
Db	2859	GATGTACTTAATGAGCTTCCTCATCTTCATTTGTAATAATTAATTTGATGTATGATCACCATT
Qy	9	GATGTACTTAATGAGCTTCCTCATCTTCATTTGTAATAATTAATTTGATGTATGATCACCATT
Db	2919	GGTCTCTCTCCCTCCCGTTTGTGTAATAATATCAGATATGCACTCCGAGGCCACTTTGGTCT
Qy	69	GGTCTCTCTCCCTCCCGTTTGTGTAATAATATCAGATATGCACTCCGAGGCCACTTTGGTCT
Db	2979	TCAGGTATGATATCCCTATTAATCTATCTGAAAGGAAAAATAGAGCCAAAGCTCTGGTCTCA
Qy	129	TCAGGTATGATATCCCTATTAATCTATCTGAAAGGAAAAATAGAGCCAAAGCTCTGGTCTCA
Db	3039	AATATATAGAAATATGCTTCTTCTTATGCTTCAGAGCTATGTGTGTAATAACAATAGAGGT
Qy	189	AATATATAGAAATATGCTTCTTCTTATGCTTCAGAGCTATGTGTGTAATAACAATAGAGGT
Db	3099	CTAATCTCTTGAAGATGAGGGGCTTTTATCTCTTAAAGAAATATGTCCCCAGATTATTTG
Qy	249	CTAATCTCTTGAAGATGAGGGGCTTTTATCTCTTAAAGAAATATGTCCCCAGATTATTTG
Db	3159	CACCTTTAGAGAGAAAGCCCAAGATATGTA-GAGGTGTGAGGTGGGCCATCAGTGGAGCAC
Qy	308	CACCTTTAGAGAGAAAGCCCAAGATATGTAAGGGGTGTGTGCTGGCCATCAGTGGAGCAC
Db	3218	GAAGAGAAATGGGATATCATTTGTGGGAGAGAAAGAAAGTTTCTCAGAGGCTCTCCACT
Qy	368	GAAGAGAAATGGGATATCATTTGTGGGAGAGAAAGAAAGTTTCTCAGAGGCTCTCCACT
Db	3278	GCTAAAGTTTTTTGTGAGATGTGTATCTGTGCTTCCTGGATTTTGACTTTTAAAGGAATTA
Qy	428	GCTAAAGTTTTTTGTGAGATGTGTATCTGTGCTTCCTGGATTTTGACTTTTAAAGGAATTA
Db	3338	TTCTGGCAGCAGATGATATTTCTTGGATGATCTTGCTCTTATTTCTCCTTTTGTGT
Qy	488	TTCTGGCAGCAGATGATATTTCTTGGATGATCTTGCTCTTATTTCTCCTTTTGTGT
Db	3398	GTGTGTGTGTGTGTGTGTGCTATATGCTTTCATTTGTAACCTCATCTGCTTATGAGAGT
Qy	548	GTGTGTGTGTGTGTGTGTGCTATATGCTTTCATTTGTAACCTCATCTGCTTATGAGAGT
Db	3458	GGGCTCTCTAATAAGGAACTGTGTAACCTTCATTTGAGAGAGGATGTATGAGAGAAATA
Qy	608	GGGCTCTCTAATAAGGAACTGTGTAACCTTCATTTGAGAGAGGATGTATGAGAGAAATA
Db	3518	GGACTTAATTCACATAGAGGCTCTCATCTCACACCTTAAAGAGAGATTTCTAGAAAAAC
Qy	668	GGACTTAATTCACATAGAGGCTCTCATCTCACACCTTAAAGAGAGATTTCTAGAAAAAC
Db	3578	TGGGCGAATTTTCTTTGTTCTCATCATTTTAAATGTGGCAGGCTGTTCAGTTTCTTAC
Qy	728	TGGGCGAATTTTCTTTGTTCTCATCATTTTAAATGTGGCAGGCTGTTCAGTTTCTTAC

```

QY      3638 TCTTACCTATGATATTTCTTCTGTAAGCTGTCCAAAAAGAAAAAGACCCTACGTGT 3697
      |||
Db      788 TCTTACCTATGATATTTCTTCTGTAAGCTGTCCAAAAAGAAAAAGACCCTACGTGT 847
QY      3698 CTCTTGACCTTTCTTTCTTTGATCCCTCAGTTCTTCTTTGATTTGAGCATGTGCGGTCC 3757
      |||
Db      848 CTCTTGACCTTTCTTTCTTTGATCCCTCAGTTCTTCTTTGATTTGAGCATGTGCGGTCC 907
QY      3758 TAATTTGGGATAGATTAGCAAAATTTAAACAATGTGTGTGTGTCCTTACCAGGGAGCTC 3817
      |||
Db      908 TAATTTGGGATAGATTAGCAAAATTTAAACAATGTGTGTGTGTCCTTACCAGGGAGCTC 967
QY      3818 CCCAGTTTCTGACTTGAATAGACTGAGAAAGATCCACGAGGTGCTATCTGCGCAATTT 3877
      |||
Db      968 CCCAGTTTCTGACTTGAATAGACTGAGAAAGATCCACGAGGTGCTATCTGCGCAATTT 1027
QY      3878 AAGTAGATTCTATTTCTTGGTTCCTCCCTCCCTGAGGACCTCTATTTTATTTGCCCC 3937
      |||
Db      1028 AAGTAGATTCTATTTCTTGGTTCCTCCCTCCCTGAGGACCTCTATTTTATTTGCCCC 1087
QY      3938 TCTTCTAGGTTAATTCCTCTTGAATTTGAATTTGTGAGAGAGAGTTGAGACATGATTT 3997
      |||
Db      1088 TCTTCTAGGTTAATTCCTCTTGAATTTGAATTTGTGAGAGAGAGTTGAGACATGATTT 1147
QY      3998 AGCAAGTTCCAAAGTCCAAATTTAACAAGTGTGAGAGTGTGGGGGAAATTTGCTTTAT 4057
      |||
Db      1148 AGCAAGTTCCAAAGTCCAAATTTAACAAGTGTGAGAGTGTGGGGGAAATTTGCTTTAT 1207
QY      4058 TTTTCCCTCATGAGGATACAAACACTGTAATTCATCTTCAACTGAAAGGCCCTGCACTTC 4117
      |||
Db      1208 TTTTCCCTCATGAGGATACAAACACTGTAATTCATCTTCAACTGAAAGGCCCTGCACTTC 1267
QY      4118 TCCATAAACATAGTTGTTGTTTCTTTTAAACAAGTTTAAAGCTAGTTTAAATTA 4177
      |||
Db      1268 TCCATAAACATAGTTGTTGTTTCTTTTAAACAAGTTTAAAGCTAGTTTAAATTA 1327
QY      4178 AAAAAATTCCTGTCGTCTACTCTGCTTTGTTTATGCCCATTTCAATTTGTTGCTG 4237
      |||
Db      1328 AAAAAATTCCTGTCGTCTACTCTGCTTTGTTTATGCCCATTTCAATTTGTTGCTG 1387
QY      4238 TGTGTAATTCATTAATTTGATGATTCATTTGATGATTTGATGATTTGATGATTT 4297
      |||
Db      1388 TGTGTAATTCATTAATTTGATGATTCATTTGATGATTTGATGATTTGATGATTT 1447
QY      4298 ATCTTATAAAGATTCATTTGTAATTAATTAATTAATTTGATGATTTGATGATTT 4357
      |||
Db      1448 ATCTTATAAAGATTCATTTGTAATTAATTAATTAATTTGATGATTTGATGATTT 1507
QY      4358 AA 4359
      ||
Db      1508 AA 1509

RESULT 10
BC043039      EC043039      2095 bp      mRNA      linear      ROD 21-OCT-2003
DEFINITION   Mus musculus RIKEN cDNA 4121402D02 gene, mRNA (cDNA clone MGC:57934
IMAGE:5701204), complete cds.
ACCESSION   BC043039
VERSION      BC043039.1 GI:27696670
KEYWORDS     MGC.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (baes 1 to 2095)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heish,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Lomg,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
12477932
2 (baes 1 to 2095)
Strausberg,R.
Direct Submission
Submitted (09-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas L. Casavant.
Web site: http://genome.uiowa.edu; tom-casavant@uiowa.edu
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
Fisler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,
Casavant,T., Soares,M.B.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21311987.
FEATURES
source
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:57934 IMAGE:5701204"
/issue_type="Brain, mouse 15.5 dpc"
/clone_lib="NIH BMP_EVO"
/lab_host="DH10B"
/notes="Vector: pYX-ASC"
1..2095
/gene="4121402D02R1k"
/notes="synonym: 4930463F05R1k, 2810017F12R1k"
/db_xref="LOCUSID:74026"
/db_xref="MGI:1921276"
41..1432
/gene="4121402D02R1k"
/codon_start=1
/product="4121402D02R1k protein"
/protein_id="AAH41039.1"
/db_xref="GI:27696671"
/db_xref="MGI:1921276"
/db_xref="LOCUSID:74026"
/db_xref="MGI:1921276"
/translation="MTMSAVPKAAAPAGNPNBQRLDYRAAALGPPDESGAERH
FLPRRLKLEPPPLASGSGSPSPACGKGGLLPAGAAAGQOEGSGSV
LPCPPATKQAGTIGEPVAGSGSPRYQAVLPYONGSIVAAKKEPTPMAGDGS
AAPPAATADPAGPPPLPAGPPPLAFTATATLAASBRKWSIRKPGGGGSGAS
SQACIKQILLILIEQQOOLQKEKEIEBLSERDTLARIERMRMQLVARD

```

ORIGIN

NEKERHKLQGYTPERBEAELESEKI KLEFOPELCTEQALPESKPSGSGKHKK
TPGNTERTKTPVKRLAPESFSGVKTPTKPSHPIKEEPGSISETCKRELNSQETPEK
RSSVDTPPRLSTPOKSGTSHPEKKAESSEMEDPLYSTETMYLCKRMHQPPSPDLJRE
SSPKEETVASKA"

Query Match 29.3%; Score 1281.2; DB 10; Length 2095;
Best Local Similarity 94.8%; Pred. No. 1,2e-234;
Matches 1348; Conservative 0; Mismatches 68; Indels 6; Gaps 2;

QY 38 CCCTCCCGGCTCGGNGCCCGGCGCTGCCGACCACTATGACCTAGATCCGGGT 97
DB 1 CCCTCCCGGCTCGGNGCCCGGCGCTGCCGACCACTATGACCTAGATCCGGGT 60
QY 98 GTTCAAGGCGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 157
DB 61 GTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
QY 158 GGGTGGCGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 217
DB 121 GGGTGGCGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
QY 218 CCGGCACTGTAAGCTCAAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 277
DB 181 CCGGCACTGTAAGCTCAAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
QY 278 CCGGCGCTCCCGGCGGCGGCT---GCGGCGGCAAGGCGGCGGCGGCGGCGG 334
DB 241 CTGGGCTCTCCAGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
QY 335 GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 394
DB 301 GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
QY 395 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 454
DB 361 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
QY 455 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 511
DB 421 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
QY 512 GGGCAAGAGCTCAAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571
DB 481 GGGCAAGAGCTCAAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
QY 572 CGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631
DB 541 CGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
QY 632 CACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 691
DB 601 CACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
QY 692 CCCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 751
DB 661 CCCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
QY 752 CTTCTGCTGCAATTGGAAGCTCAAGAGCGGCGGCGGCGGCGGCGGCGGCGG 811
DB 721 CTTCTGCTGCAATTGGAAGCTCAAGAGCGGCGGCGGCGGCGGCGGCGGCGG 780
QY 812 GGGAGTCGAGAGAGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 871
DB 781 GGGAGTCGAGAGAGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
QY 872 AAGCGGATGAGAGAGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 931
DB 841 AAGCGGATGAGAGAGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
QY 932 CTATGAAGCTGAG 991
DB 901 CTATGAAGCTGAG 960

QY 992 GCCGAGCTTTCCGAGACATCCAGACTGCTGCTGCCAAGCCCTTCATGTGGCGGAG 1051
DB 961 GCCGAGCTTTGGAGAGCTCCAGAGCTGCTGCCAAGCCCTTCATGTGGCGGAG 1020
QY 1052 TGGAAAGGACATAAAGAAATCCCATTTGGAAGTACGAAAGAAAGACTCTGTTAA 1111
DB 1021 TGGAAAGGACATAAAGAAATCCCATTTGGAAGTACGAAAGAAAGACTCTGTTAA 1080
QY 1112 AAGCTGCTCTGAAATTTTCAAAAGTCAAAACAAAACCTTAAGCACTCTCTATTAA 1171
DB 1081 AAGCTGCTCTGAAATTTTCAAAAGTCAAAACAAAACCTTAAGCACTCTCTATTAA 1140
QY 1172 AGAGAAACCTTGCTGCTTCTATCTGAACCTGTTTGAACGTGAATTGAGAGCCAAAG 1231
DB 1141 AGAGAAACCTTGCTGCTTCTATCTGAACCTGTTTGAACGTGAATTGAGAGCCAAAG 1200
QY 1232 AACCCAGAAAGAGCCCGGCTTCAAGTGAACACCCCAAGACTCTCCCAAAA 1291
DB 1201 AACCCAGAAAGAGCCCGGCTTCAAGTGAACACCCCAAGACTCTCCCAAAA 1260
QY 1292 GGGAGCCAGACCCATCCCAAGAGAAAGGCTTCTCAAGTGAATAGAGATTGGCCGTA 1351
DB 1261 GGGAGCCAGACCCATCCCAAGAGAAAGGCTTCTCAAGTGAATAGAGATTGGCCGTA 1320
QY 1352 CTTTCCACACAGAAATGATTTGCTGCTTGGACACAGCTCCCATCAAGCTTACC 1411
DB 1321 CTTTCCACACAGAAATGATTTGCTGCTTGGACACAGCTCCCATCAAGCTTACC 1380
QY 1412 ATTAAGGAAATCTCTCCAAAGAGAGAGAGACTGTAGCAAG 1453
DB 1381 ATTAAGGAAATCTCTCCAAAGAGAGAGAGACTGTAGCAAG 1422

RESULT 11
BC055715
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE

BC055715 2087 bp mRNA linear ROD 21-OCT-2003
MUS musculus RIKEN cDNA 4121402D02 gene, mRNA (cDNA clone MGC:66782
IMAGE:6414174), complete cds.
BC055715 GI:33585901
MGC.
Mus musculus
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2087)
Straussberg, R.L., Feingold, E.A., Grouse, J.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, U.,
Scheleton, M., Soares, M.B., Bonaldo, M.F., Cassavani, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Aramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Sánchez, A., Whitting, M., Madan, A., Young, A.C., Shvachenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E.,
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 2087)
Straussberg, R.
Direct Submission

JOURNAL Submitted (01-AUG-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk
Email: cgabs-rc@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

FEATURES
source
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRM Plate: 125 Row: m Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21311987.
Location/Qualifiers
1..2087
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:66782 IMAGE:6414174"
/issue_type="Brain, pool of enriched mouse brain tissues from 13.5, 14.5, 16.5 and 17.5 dpc"
/clone_11b="NIH_BMAP_FR0"
/lab_host="DH10B"
/note="Vector: pYX-ASC"
1..2087
/gene="4121402D02R1k"
/note="synonyms: 4930463F05R1k, 2810017P12R1k"
/db_xref="LocusID:74026"
/db_xref="MGI:1921276"
15..1406
/gene="4121402D02R1k"
/codon_start=1
/product="4121402D02R1k protein"
/protein_id="AAH5715.1"
/db_xref="GI:33585902"
/db_xref="LocusID:74026"
/translation="MTMSRAVFKAAAPAGNPNRDLVYRAALGPPENSGAARAH PLPRHKLKEPPLASSGGSPSPAGCGGSKGGLIPAKAARGGQDESGSYV LCPPPATKQAGIGEPVAAAGCSPRPKIQAVLP1QTSIVAAKEPTPMAGDGAG AAPATVADDPAGPPLPLPGPPPLATATATLASBGRWKSIRSPLOGGGGAS SQAACTKQILLQDLIEQQOOLQAKKEKIEBKESBDTLAIRSRMERKMLVARD NEKERHKLQGYEPEREBEALSEKIKLEBPCLCETSOALPSPKSCGSGGHRK TPENERTKTPVKKLAPEFSKVTXTPKPSPIKEPCGSISETVCKRELRSDTPKRP RSVVDTPPRUSTPQKPSHPKXKAPSSMEDLPYLTMTYLCRWHPSPPLPIRE SSPKKEETVASKA"

ORIGIN
Query Match 28.8%; Score 1260; DB 10; Length 2087;
Best Local Similarity 94.9%; Pred. No. 1.3e-230;
Matches 1325; Conservative 0; Mismatches 65; Indels 6; Gaps 2;

QY 64 TGCTCGGACCACTATGACCATGATCCCGGTGTTCAGAGCGCGCCCGCTTCGCG 123
DB 1 TGCTCGGACCACTATGACCATGATCCGAGTGTTCAGAGCGCGCCCGCTTCGCG 60
QY 124 GCGGCAATCTGAGCAGCACTGACTAGAGCGGCTGCGGCTGGCGGCGCCGAGG 183
DB 61 GCGGCAACCCCGGAGCGAGCTGACTACGAGCGGCTGCGGCTGGCGGCGCCGAGG 120
QY 184 AGGAGCTGGGGGCGGAGGCCACTTCTCCCGGCGACCGTAAGCTCAAGAGCGCG 243
DB 121 ACGAGTCGCGGCGGCGAGGCCATTCTCCCGGCGACTCGTAGCTCAAGAGCGCGG 180

QY 244 GGGCCCCGCTGGCTCTCTCCACAGGCGGAGAGCCCGCGCTTCCCGCGCGCT---GGG 300
DB 181 GGGCCCCGCTGGCTCTCTCCACAGGCGGAGAGCCCGCTCTCTCCACGCGCTGGCGG 240
QY 301 GCGGCAAGGCGCGGGCTGTACTCTCCGCGGAGGCGCGCCCGGCGACAGAGAGAG 360
DB 241 GCGGCAAGGCGCGGGCTGTACTCTCCGCGGAGGCGCGCCCGGCGACAGAGAGAG 300
QY 361 GCTGGGCGGATTCGATGCCCTTTCCTGTTCGCGCCCGGCGACCAAGCAAGCGGATTTG 420
DB 301 GCTGGGCGGATTCGATGCCCTTTCCTGTTCGCGCCCGGCGACCTACTAACAAGCGGATTCG 360
QY 421 GGGGGAGCGCTGCGGAGCGCGAGCCGCTGACAGCCCCCGGCGCAAGTATCAGGCGCTGC 480
DB 361 GCGGGAGCGCGATTCGAGCGCGCGCTGCTGACAGCCCCCGGCGCAAGTATCAGGCGCTGC 420
QY 481 TGCCCATTCAGACGAGCTCTCTC---GTGCGGCGGCGCAAAAGAGCTTACGCCCTGGGCTG 537
DB 421 TGCCCATTCAGACGAGCTCTCTATCTGTGGCGGCGCGCAAAAGAGCTTACGCCCTGGGCTG 480
QY 538 GGGACAAAGGATGGGCGGCGCTCCCGCTGCGACCGCTCGAGACCGGCGGAGACCCCGAC 597
DB 481 GGGACAAAGGATGGGCGGCGCTCCCGCTGCGACCGCTCGAGACCGGCGGAGACCCCGAC 540
QY 598 CACTACCTTGCCTCGGCGCGCCACCCCTTGGCGCCACCGCCACCGCGGAGCCTTGGCGG 657
DB 541 CACTACCTTGCCTCGGCGCGCCACCCCTTGGCGCCACCGCCACCTGTGGAGCCTTGGCGG 600
QY 658 CCAGCGAGGCGAGATGGAAGAGATGAGGAGAGCCCTCGGCGGATGGTGGCGGCTCGG 717
DB 601 CCAGCGAGGCGAGATGGAAGAGATGAGGAGAGCCCTCTCGGCGGATGGCGGCGGCTCGG 660
QY 718 GAGCCTCAATCAGGCGCGCTGCTCAACAGATCTTGTCTGTCAATTGACCTCATCG 777
DB 661 GAGCCTCAATCAGGCGCGCTGCTCAACAGATCTTGTCTGTCAATTGACCTCATCG 720
QY 778 AACAGCAGACGACGACCTGCGAGGCCAAGAGAAAGAGATCGAGAGCTGAAATCGAGAG 837
DB 721 AACAGCAGACGACGACCTGCGAGGCCAAGAGAGAGATGAGGAGCTGAAATCGAGAG 780
QY 838 GAGACAGCGCTCTGCTCGGATTTGAACGTATGGAAGGCGGATGACGCTGTAAAGAG 897
DB 781 GAGATAGCGCTCTGCTCGGATTTGAACGTATGGAAGGCGGATGACGCTGTAAAGAG 840
QY 898 ATTAACGAGAAAGAGAGCAAGCTGTTTCAGGCTATGAACTGGAAGAGAGAGAGAA 957
DB 841 ATTAACGAGAAAGAGAGAGCAAGCTGTTTCAGGCTATGAACTGGAAGAGAGAGAG 900
QY 958 CAGAGCTATCTGAGAAATTTAACTGAGTGCAGCGGAGCTTTCGAGACATCCGAGA 1017
DB 901 CAGAGTGTCTGAGAAATTTAAATTTGAGACGCGCAGCGGAGCTTTCGAGACATCCGAG 960
QY 1018 CTCTGCTCCCAAGCCCTTCTCATGTGGGCGGAGTGGAAAGGAGCACTAAAGGAATCCC 1077
DB 961 CTCTGCTCCCAAGCCCTTCTCATGTGGGCGGAGTGGAAAGGAGCACTAAAGGAATCCC 1020
QY 1078 CATTTGAAATCAGAGAAAGAGACTCTGTTTAAAGCTGCTCTGTAATTTTCAAAAG 1137
DB 1021 CATTTGAAATCAGAGAAAGAGACTCTGTTTAAAGCTGCTCTGTAATTTTCAAAAG 1080
QY 1138 TCAAAAAGAAAGCTCTTAAGCACTCTCTATTAAAGAGAAACCTGTGTCTTATCTTG 1197
DB 1081 TCAAAAAGAAAGCTCTTAAGCACTCTCTATTAAAGAGAAACCTGTGTCTTATCTTG 1140
QY 1198 AAACCTTTTGTAAAGCTGAATTTGAGAGGCAAGAAACCCCAAGAAAGCCCGGTCTTTCAG 1257
DB 1141 AAACCTTTTGTAAAGCTGAATTTGAGAGGCAAGAAACCCCAAGAAAGCCCGGTCTTTCAG 1200
QY 1258 TGGACACCCCAACCAAGACTCTCACTCCCAAAAGGAGCCAGACCCATCCCAAGAGAG 1317
DB 1201 TGGATATCCCAACCAAGACTCTCACTCCCAAAAGGAGCCAGACCCCAAGAGAGAG 1260

QY 1318 AACCTTTCAGTAGAGATGAGAAATTTGCCGTACCTTTCCACACAGAAATGTATTTGT 1377
DB 1261 AAGCCTTTCAGATGAGATGAGAAATTTGCCGTACCTTTCCACACAGAAATGTATTTGT 1320
QY 1378 GTGCTTGAGACGACGCTCCGCCATCAACCTTACATTACGGGAATCTCTCCAAAGAG 1437
DB 1321 GTGCTTGAGACGACGCTCCGCCATCAACCTTACATTACGGGAATCTCTCCAAAGAG 1380
QY 1438 AGAGAGCTGAGCAAG 1453
DB 1381 AGAGAGCTGAGCAAG 1396
RESULT 12
AC119462.3
WPCOMMENT
Sequence split into 4 fragments LOCUS AC119462 Accession AC119462
Fragment Name Begin End
AC119462.0 1 110000
AC119462.1 100001 210000
AC119462.2 200001 310000
AC119462.3 300001 406773
Continuation (4 of 4) of AC119462 from base 300001 (AC119462 Rattus norvegicus clone CH2)
Query Match 19.1%; Score 835.8; DB 2; Length 106773;
Best Local Similarity 69.2%; Pred. No. 4e-149;
Matches 1801; Conservative 0; Mismatches 537; Indels 263; Gaps 38;
QY 1832 GAATTTTACGCTACCTGCTGTGATGAGCGGATGAGATTTGAGATTCAGAGAA 1891
DB 33404 GAATTTTACGCTACCTGCTGTGATGAGCGGATGAGATTTGAGATTCAGAGAA 33463
QY 1892 GCAAAACCTCCACGAGCGTGTGAAATAGCTGTGTGCGACGAACCTGTCTTCAGAT 1951
DB 33464 GCAAAACCTCCACGAGCGTGTGAAATAGCTGTGTGCGACGAACCTGTCTTCAGAT 33522
QY 1952 AGTTTATGACATGACATTC-CGAGAGTGAGAGAGACCTGTATATGATCCTTCTCA 2010
DB 33523 AGTTTATGACATGACATTC-CGAGAGTGAGAGAGACCTGTATATGATCCTTCTCA 33582
QY 2011 CATATGTTATCTGCTGTATATACCTTTTCACTTC- -TTGACTTTGTTTC 2064
DB 33583 TGTATGTTATCTGCTGTATATACCTTTTCACTTC- -TTGACTTTGTTTC 33642
QY 2065 ATACTCTGATTCACA- - - - -AAAACCTTTTCACTGCGCTAATGTGAT 2111
DB 33643 TTTACTCTGATTCACAAAACAAAACAAAGAAACCTGTGCTATGAGCTAATGTGAT 33702
QY 2112 ATGAGAGGATTTGGGATTTCTTTTCCCTTTTGGGAATGGGCTCTCAAGCTAAAGT 2171
DB 33703 ATGAGATTTGGAATTTCTTTTCCCTTTTGGGAATGGGCTCTCAAGCTAATGT 33759
QY 2172 ATAGATGAGATTCAGAAATTTGAGAGGCTGTGTTCTATATGATTTGCTATGTTAAG 2231
DB 33760 ATAGATGAGATTTGAGAGGCTGTGTTCTATATGATTTGCTATGTTAAG 33816
QY 2232 GGGTAAAGGGCTCTCTTCAATGAGCACTGTGGAAGATGAGAGAGAGAGAGAGAGAG 2287
DB 33817 GGGTAAAGGG- - -CTCTTCAATGAGCACTGTGGAAGATGAGAGAGAGAGAGAG 33873
QY 2288 GAGCTGCTGCTGATGAGCACTTTCTCAACCTGTGACACCTCTTATATGAGGATAGT 2347
DB 33874 GAGCTATGCTGCTGATGAGCACTTTCTCAACCTGTGACACCTCTTATATGAGGATAGT 33933
QY 2348 GATTTTAAACCTTAAATTAACAAACAAACCTCAACATGAGCTTTAGAGACAGAGAGA 2407
DB 33934 GATTTTAAACCTTAAATTAACAAACAAACCTCAACATGAGCTTTAGAGACAGAGAGA 33959
QY 2408 ATAGCAAGTGAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2467
DB 33960 - - - - -TGAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 33995
QY 2468 GGTATGATTAACCTGTGAAGAGATGTTGTTCAATTTGAAACTAATTTCTGTATACAGTATG 2527

DB 33996 - - - - -GCTGGGCACTGGCATG 34012
QY 2528 TGGAGAGGATGTTGGCTGTGATTAATTTTCACTTAATGAGTAACAATTTCTTACTG 2587
DB 34013 TGGAGAGGATGTTGGCTGTGATTAATTTTCACTTAATGAGTAACAATTTCTTACTG 34071
QY 2588 CTCAAAACCAAAATCTTTGAAAAAGAGTGGGGATGTTAGTTTCAGAAACAATTACA 2647
DB 34072 ACCCCACAGAGAAATGTTTGAAGAGGGGTTGGGGGTTAAGTTCTG- - - - - 34120
QY 2648 GCTGTAAACAAAGCACTTAATTTTGGATGAGCAATGCCAAACCTGTATTAATGTCTT 2707
DB 34121 - - - - -AAGCATTCAGTTGTCACTTGTAGGATGAGACACAGAGCTGTGTGAGATGTCTCT 34175
QY 2708 GTATCATCATCTTCAAGTATTCCTTCACTTGGGGCTCACTCTTTAGAGAGAGCTCTG 2767
DB 34176 GTCCATGTCTCTCAAGTATTCCTTGTGGGGCTTCAAGCTTATTTGGCCAACTTTTG 34235
QY 2768 GTGTGGGATGAGACTTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2826
DB 34236 GTGTGTAGTGAAGATTTGGGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 34295
QY 2827 GAGTGGCAATGTCTATCATCTTGAACAAACAGATGTACCTATGAGCTTCTCCATTGAC 2886
DB 34296 CGCTGACAGTTGTCTATCATCTTGAACAAACAGATGTGACAGATGAACCT- - -TATTGCG 34352
QY 2887 TTTGTAAAAAT-ATTGTATGTGTACATCTTGTCTCTCCCTCCCTTTTGT- - - 2942
DB 34353 TTATGTAAGATGATGATATCATGTATCATCTTGTTCATATCTCTGCTTGTGTTCT 34412
QY 2943 - - - - -AAAAATCAGATGACA 2959
DB 34413 TACCGAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 34472
QY 2960 CTCACAGGCACTTTGATCTCACTGTATGATTCCTATTTAATCTGAAAGAAAAATAGA 3019
DB 34473 CTCACAGGCACTTTGATCTCACTGTATGATTCCTATTTAATCTGAAAGAAAAATAGA 34530
QY 3020 GCCAAGACCTGTGTCTCAATATATGAGAAATGCTTTCTTATAGTCTTACAGACTATG 3079
DB 34531 GCCAAGACCTGTGTGTCT- - -TGTATGAGAAATGCTTTGCTGAGGCTTACAGAACTGTTG 34587
QY 3080 TGTGAAGAACAGTGGGCTATATCTCTTGAAGATGAGGGCTTTATCTTAAAGAGAA 3139
DB 34588 TATTAAGAACAGTGGGCTATATCTCTTGAAGATGAGGGCTTTATCTTAAAGAGAA 34646
QY 3140 TATGTCCCAAGATTTATGACATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3199
DB 34647 CATGTCCCGGATTTATGACATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 34706
QY 3200 GGGCCATCATGTGAG 3259
DB 34707 GGGCCATCATGTGAG 34764
QY 3260 CCTCAGGGGCTGCCCTGTAAAGTTTGTGAGATGTGATGTGCTGCTGTGAT 3319
DB 34765 TCTCA-GGGCTGCCCTGTAAAGCTTTT-TGAGATGT-ATCTGTGCTCCTGAGTT 34821
QY 3320 TGACTTTAAAGAAATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3379
DB 34822 TGACTTTAAAGAAATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 34881
QY 3380 TTAATTTCTCTTTTGT 3439
DB 34882 TTAATTTCTCTT- - - - -TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 34934
QY 3440 CCATCTGCTTGAAG 3499
DB 34935 CCATCTGCTTGAAG 34994
QY 3500 AGAGATGAGAGAGAAATGAGACTTAATTTCCATGAGGGGCTCTCATCTCAACCTTAAAGA 3559


```

Db      34995 AGGATGTAGAGAGAAAAGAGACTTCCTCCTAGTGTGGCTTCATCTCACACCTTAAGGA 35054
Qy      3560 ---GGAGATTCTAGAGAAAACGTGGCCAGATTTCTTTGTCTCCATCATTTAAATGTG 3516
Db      35055 TGGGGATATATCCAGAAAGTTGGGGCAGATTTTTT-----TTTAAATGTG 35102
Qy      3617 CAGGCGTTCAGTTTCTTACTCTTACTCTTATGTATGTATTTCTTGTAAGTG--TCCAA 3674
Db      35103 CCCACGTCTTATAGTCTCTTACTCTTATGTATGTATTTCTTATTAAGCTCTCTCAA 35162
Qy      3675 AAGAAAAAGACCCTATCAGTGTCTCTTGAATTTGTTTGAATCC--CTCAGTTTCTTC 3732
Db      35163 AATATATACCTCAATCCCGTGTCTCTTAACTTTGTTTGTGATCCCTCAGTTTCTTC 35222
Qy      3733 TTGATTTCAGCATGTGTGGGTTCTTAATTTGGGATAGATTAGCAAAATTAACATTG 3792
Db      35223 TTGACACACAGATTTTGTCAATTTCCCTAA--TTTGGATATGGGTAAATAGACTTAACGCCCA 35281
Qy      3793 TGTGTGTGCTTATCCAGGGGACTCCCGAGTTTCTGACTTGAATGACTGAGAGAAATC 3852
Db      35282 GGTTCCTGCTTATTTAGAGAGAGTCTCATATCTCTTTTTCG--TAGACTGAGAGAAATC 35340
Qy      3853 CAGAGAGCTCTATCTTGTGGCCAGATTTAAGATTTCTATTTCTTGTGTTCCCTCCCT 3912
Db      35341 CATGAGTGTCTATCTGCGACAGACTGAAGTGAATTTCTATTGCTTTGGTTC--CTTTCCCT 35399
Qy      3913 GAGAGACTCTATTTATTTGTCCCTCTTCTAGATTAATTCCTCTTGAATTTGACTTTGT 3972
Db      35400 GAGGACCTTATATTTATTTATTTATGTCCTCTAG-----CTGCTTTTGGTTTGAAGATGC 35454
Qy      3973 TGAGAGAGAGGTTGACAGTATGCAAAAGTTCCAAAGTCAAAATTAACAGTGTGTAG 4032
Db      35455 TGAAGTGTGTTGAGAGAGAGTGTGAGCAAAAGTTCCAGAGCAAGACAGATGTGT--- 35511
Qy      4033 AGTGTGGGGGAAAATTAAGTCTTATTTTCCCTACATGAGGATCAACACTGTG---AAT 4089
Db      35512 GAGGTGTGTTGGGGAATTAATTTATTTTCCCTACGTTGGGATCAACACTGTGTGAGATT 35571
Qy      4090 CAATCTTCAACTGA-----AGCCCTGCAATTTCTCTTAAACATAGTT--GTTGTTTTC 4143
Db      35572 CAGCCTTCAACAGAAAAGAGAGGCCCTGCAATTTCTCTTAAACATAGTTCTTCTCTTC 35631
Qy      4144 TTTAACAAAGTTTAACTAGTGTATTAATTAATTA---AAAAATGCTGTCTCTACTT 4201
Db      35632 TTTAACAAAGTTTAACTAGTGTATTAATTAATTAAGAGAAAGTGTCTCTCTCAACT 35691
Qy      4202 CAGCTTGTTTTATGCCCATTTCATATTTGTGTCTGTGTGTAATTCATAACTTTGATA 4261
Db      35692 CAGCTTGTTTTATGCTGTTCATATTCGTGTGTGTGTGTAATTCAT--ACTTTTGATA 35750
Qy      4262 CCATTTCTGATGTGTAATTAATTTGCTGTCTGTAATTAATTAAGATTCATTTGATA 4321
Db      35751 CCATTTCTGATGTGTAATTAATTTGCTGTCTGTAATTAATTAAGATTCATTTGATA 35810
Qy      4322 ATTAATTAATTTGCTGTGTTAA 4342
Db      35811 ATTAATTAATTTGCTGTGTTAA 35831

```

AUTHORS

Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benham, F., Biswal, O. K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cre, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Decker, D., Delgado, O., Denson, S., Derramo, C., Ding, Y., Din, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgeson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hune, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, U., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensbawa, L., Louised, H., Lozada, R. J., Lu, X., Ma, Y., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, B., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, S., Milsavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muniadas, M., Murphy, M., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokeleleh, O., Okwunu, G., Olarnunsegon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Frankoch, C., Pioppier, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reich, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Snead, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, A., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, L., Trejos, Z., Umanu, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, D., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

REFERENCE

2 (bases 1 to 183618)

REFERENCE

Rat Genome Sequencing Consortium.

REFERENCE

Submitted (21-Aug-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

3 (bases 1 to 183618)

REFERENCE

Rat Genome Sequencing Consortium.

REFERENCE

Submitted (18-Sep-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

On Sep 18, 2002 this sequence version replaced gi:22380622.

REFERENCE

The sequence in this assembly is a combination of BAC based reads

REFERENCE

and whole genome shotgun sequencing reads assembled using Atlas

REFERENCE

(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the

REFERENCE

sequence may extend beyond the ends of the clone and there may be

REFERENCE

contigs that consist entirely of whole genome shotgun sequence

REFERENCE

reads. Both end sequences and whole genome shotgun sequence only

REFERENCE

contigs will be indicated in the feature table.

Center: Baylor College of Medicine

Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: G001
Center clone name: CH230-278A19

----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 176229 bases at least Q40
Consensus quality: 178023 bases at least Q30
Consensus quality: 179050 bases at least Q20
Estimated insert size: 200566; sum-of-coverage
Quality coverage: 4x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 35058: contig of 35058 bp in length
* 35059 35158: gap of unknown length
* 35159 139365: contig of 104207 bp in length
* 139366 139465: gap of unknown length
* 139466 140917: contig of 1452 bp in length
* 140918 141017: gap of unknown length
* 141018 142330: contig of 1213 bp in length
* 142331 145121: contig of 2791 bp in length
* 145122 145221: gap of unknown length
* 145222 148094: contig of 2873 bp in length
* 148095 148194: gap of unknown length
* 148195 183618: contig of 35424 bp in length.
* location/Qualifiers
1. 183618
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-278A19"
1. 1263
/note="wgs_contig"

ORIGIN
misc_feature
Query Match 19.1%; Score 835.8; DB 2; Length 183618;
Best Local Similarity 69.2%; Pred. No. 4.3e-149;
Matches 1801; Conservative 0; Mismatches 537; Indels 263; Gaps 38;

QY 1832 GAATTTGAGTACCTCGTGTGATGAGCGTACGATGAGATGAGATCCAGAA 1891
Db 57761 GAATTTGAGTACCTCGTGTGATGAGCGTACGATGAGATGAGATCCAGAA 57702
QY 1892 GCAAAACCTCACCGGCTGTGAGAGATGAGATGAGATGAGATGAGATGAGAT 1951
Db 57701 GCAAAACCTCACCGGCTGTGAGAGATGAGATGAGATGAGATGAGATGAGAT 57643
QY 1952 AGTTGTAGACGCGATTC-C-GAGAGTGGAGAGACCTGTATATGATGATGAT 2010
Db 57642 AGTTGTAGACGCGATTC-C-GAGAGTGGAGAGACCTGTATATGATGATGAT 57583
QY 2011 CATATGTATACCTCGTGTATATACCTTTTCACTTCC-----TTGACTTGT 2064
Db 57582 TGATATGTATACCTCGTGTATATACCTTTTCACTTCC-----TTGACTTGT 57523
QY 2065 ATTACTGTATTCACA-----AAAACTTTTCACTTGGCTAATTTGATGAT 2111
Db 57522 TTATCTGTATTCACA-----AAAACTTTTCACTTGGCTAATTTGATGAT 57463
QY 2112 ATGAGAGGTATGAGATTTCTTTTCCCTTTTGGAGAAATGGGCTCTCAAGCT 2171
|||||

Db 57462 ATGAGATTT---ATTGAACTTTCTTTCCCGTTTGGGAAATGAGCTTGAAGCTCATCT 57406
QY 2172 ATAGATGAGATGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 2231
Db 57405 ATAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 57349
QY 2232 GGGTAAAGGGCTCTCTTCAATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 2287
Db 57348 GGGTAAAGGG---CTCTTCAATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 57292
QY 2288 GAGCTGTGCTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 2347
Db 57291 GAGCTGTGCTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 57232
QY 2348 GATTTTAAACCTTAAATTAAGAAACCAACCTGATGAGATGAGATGAGATGAGATGAGAT 2407
Db 57231 GTAGATTTCACTTAAAGGATTAACA----- 57206
QY 2408 ATGACAAGTGAAGCGATGAAGCAAGCCATCTTCAAGATGAAGAAAGACATGGAAGATT 2467
Db 57205 -----TGAAGATTTGAAGCAAGACACTCTCAAGAGAAAGAA----- 57170
QY 2468 GATGATTAAGCTGTGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2527
Db 57169 -----GCTGGAGACATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 57153
QY 2528 TGGGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2587
Db 57152 TGGGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 57094
QY 2588 CTCAAAACCAAAATCTTTGAGAAAGAGAGAGAGATGATGATGATGATGATGATGATGAT 2647
Db 57093 ACCCCCAAGAGAAATGATGAGAAAGAGAGAGATGATGATGATGATGATGATGATGATGAT 57045
QY 2648 GCTGTAAACAAAGCACTTATGATTTGGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 2707
Db 57044 -----AAGCATTCAGTGTCACTTGAAGATGAGATGAGATGAGATGAGATGAGATGAGAT 56990
QY 2708 GTATCATCATCTTCAAGATTTCTTCAATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 2767
Db 56989 GTATCATCATCTTCAAGATTTCTTCAATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 56930
QY 2768 GTGTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 2826
Db 56929 GTGTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 56870
QY 2827 GGTGTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 2886
Db 56869 GGTGTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 56813
QY 2887 TTTGTAAATAA-TA-ATTGTATGTGTACATCTTGTCTCTCCCTCCCTTTTGT--- 2942
Db 56812 TTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 56753
QY 2943 -----AAATATCAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2959
Db 56752 TACCGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 56693
QY 2960 CTCCTCAAGGCACTTGTGTCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3019
Db 56692 CTCCTCAAGGCACTTGTGTCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 56635
QY 3020 GCGAAGACCTGTGTCTCAATATATATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3079
Db 56634 GCGAAGACCTGTGTCTCAATATATATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 56578
QY 3080 TGTGAAACCAAGTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3139
Db 56577 TATTAACCAAGTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 56519
QY 3140 TATGTCCCAAGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3199
Db 56518 CATGTCCCGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 56459

[illegible]

QY	4262	CCATTTCGTGATGCTGTAATAATGGTCTCTCTGTGTAATAATCTTTATTAAGGCTTCAATTGTAA	4321
Db	55414	CCATTTCGTGATGCTGTAATAAGTCGGCTCTGTGTAATAATCTTTATTAAGGCTTCAATTGTAA	55355
QY	4322	ATAAACAATTGTGGCGTGTAA	4342
Db	55354	ATAAACAATTGTGGCGGTAA	55354
RESULT 14			
AL590963/c		196724 bp	DNA
LOCUS			linear
DEFINITION		Mouse DNA sequence from clone RP23-395E10 on chromosome 11,	ROD 05-APR-2002
ACCESSION		complete sequence.	
VERSION		AL590963	
KEYWORDS		AL590963.11 GI:20068458	
SOURCE		HMG.	
ORGANISM		Mus musculus (house mouse)	
REFERENCE		Mus musculus	
AUTHORS		Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
JOURNAL			
COMMENT		<p>1 Lovell, J. Direct Submission Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hamuery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On Apr 7, 2002 this sequence version replaced gi:16304739. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., paired quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWSISSPROT; Tr: TREMBL; Wp: WORMPEP; Information from the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep from the RPc1-23 Mouse PAC Library constructed by the group of Peter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBACE3.6. location/Qualifiers</p>	
FEATURES		<p>source</p> <p>1..196724</p> <p>/organism="Mus musculus"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:10090"</p> <p>/chromosome="11"</p> <p>/clone="RP23-395E10"</p> <p>/clone_1b="RPc1-23"</p>	
ORIGIN			
Query Match		18.9%;	Score 825; DB 10; Length 196724;
Best Local Similarity		69.7%;	Pred. No. 5.1e-147;
Matches 1841; Conservative		0;	Mismatches 540; Indels 262; Gaps 43;
QY	1832	GAATTTGACCTACCTGTTGGATGAGCGTATGCCGATGCAGATTGGAGATCCAGAGAA	1891
Db	53335	GAATTTGACCTACCTGTTGGATGAGCGTATGCCGATGCAGATTGGAGATCCAGAGAA	53176
QY	1692	GCAAAACCTCACCGGACGTATGGAATAATAGCGTGCAGCAAGAACCTGCTTCAGAT	1951
Db	53175	GCAAAACCTCACCGGACGTATGGAATAATAGCGTGCAGCAAGAACCTGCTTCAGAT	53117
QY	1952	AGTGTGACATGTCATTTC-CGAGATGTCAGAGACCTGTATATGTGACCTTTGTCTCA	2010

QY	4049	TAGCTTAATTTTCCCAACAGGGATACCAACACGTGGA -ATTAACTTCAACTGA -----	4103
Db	51017	T-----GCTCCCTCCCTGGGGGACCAACCTGAGACTCACTTTCAGACAGAAAAG	50966
QY	4104	-AGGCGCTGCAGTTCTCCTAAACATAGTTGTTGTT -TTCTTTAACAAGTTTAAGCT	4161
Db	50964	GAGGCCCTGCACCTCTCTGAAAAACATGTTCTTATCTTTCTTTAAACAAGTTTAAGCT	50908
QY	4162	AGTTTAAATAATAATAA -AAAAATTGCTGTGCTACTCTAGGTTGTTTATAGCC	4219
Db	50904	AGTGTTAATAATAATAAAGAAATTCCTGTCTCTCAACTCAGCTTTATTTATAGCC	50845
QY	4220	ATTTCATATTTGTGTCTGTGTGTAAATTCATATTCATTGATACATTCTGATGTAA	4279
Db	50844	GTTTCATATCGTGTGTGTGTCTGTAAATTCAT -ACTTTGATACATTTCTGATGTAA	50786
QY	4280	ATTGGTGTCTGTAATATCTTAATAAGATCAATTGAATTAACATATGTGGCTGT	4339
Db	50785	GTCGGCTGTCTGTAAATATCTTTAANAAGTTCAATTGAATTAACATATGTGGCGT	50728
QY	4340	TAA 4342	
Db	50725	TAA 50723	

RESULT 15			
BD017821/c			
LOCUS	BD017821	910 bp	DNA
DEFINITION	Novel gene and novel gene fragment cloned in human neuroblastoma.		
			linear PAT 27-AUG-2002

ORGANISM	REFERENCE	TITLE	AUTHORS	JOURNAL	COMMENT
Homo sapiens	Eukaryote: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini, Homnidae, Homo.				
	1 (bases 1 to 910)				
Nakagawara,A					
Novel gene and novel gene fragment cloned in human neuroblastoma patient: JP 2001245671-A 59 11-SEP-2001,					
CHIBA PREF,HISAMITSU PHARMACEUTICAL CO INC					
OS Homo sapiens (human)					

REFERENCE	(bases 1 to 910)
AUTHORS	Nakagawara,A.
TITLE	Novel gene and novel gene fragment cloned in human neuroblastoma
JOURNAL	Patent: JP 2001245671-A 59 11-SEP-2001; CHIBA PREF,HISAMITSU PHARMACEUTICAL CO INC
COMMENT	OS Homo sapiens (human) PN JP 2001245671-A/59 PD 11-SEP-2001 PF 07-MAR-2000 JP 2000159195 PI AKIRA NAKAGAWARA PC C12M15/09,C12Q1/68,G01N3/53,G01N3/566/(C12Q1/68,C12R1.91), CC C12M15/00 CC Novel gene and novel gene fragment cloned in human CC neuroblastoma
FH	Key
FT	source
FEATURES	location/Qualifiers
source	1..910 /organism='Homo sapiens (human)'. 1..910 location/Qualifiers

FEATURES	Location/Qualifiers
source	1..910
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
ORIGIN	

Query Match	16.5%	Score 722	DB 6	length 910
Best Local Similarity	93.1%	Pred. No. 1.3e-127		
Matches 787	Conservative	0	Mismatches 54	Indels 4
			Gaps	4

QY	Db	QY	Db
3522	TTAATTCACATGAGGGGCTCATCTCAACACCTTAAAGGAGGAGATTCTTAGAAAACTGGG	3582	CCAGATTTTCTTTGTCTTCATCATTTTAAAGTGGCAGAGCTGTTAGATTTTCTTAAGCTTT
865	TTTATCCCGTTGGGGGNTTNAINTCAACCCCTTTAGGAGGAGATTTTGGAAAAANTGNN	805	CCAGATTTTCTTTGTCTTCATCATTTTAAAGTGGCAGAGGCTCAG- TTTCTNANTCTTT
		747	

Oy	3642	ACCTATGTGATATTTCTTCGTAAAGTGTCCAAAAAGAAAAAGCCCAATGAGTCTC	3700
Db	746	ACCAATGTGA-ATTCTTCGTAAAGTGNCCAAAAAGAAAGAACCCCAATCAATGATCTCN	688
Oy	3701	TTGACTTGTGTTCTTGGATNCCCTCAGTTTCTTCTGAATTTGACGATGTGCGGATTCCTAA	3760
Db	687	TTGACTTGTGTTCTTNNATCCCTAGTTTCTTCTGATTTGACGATGTGTGCGGATTCCTAA	628
Oy	3761	TTTTGGGTAGAGTTAGCAATTTTAAACATTTGTTGTGCTTAACCAAGGGAGCTCC	3820
Db	627	TTTTGGGTAGAGTTAGCAATTTTAAACCATTTGTTGTGCTNNATCCCAAGGGAGCTCC	568
Oy	3821	AGTTCTGACTTGAAGTAGACTGAGAAAGATCCACGAGTGCTATC-TGGCCAAATTTAA	3879
Db	567	AGTTCTGACTTGAAGTAGACTGAGAAAGATCCACGAGTGCTATCTTGGCCAAATTTAA	508
Oy	3880	GTAGATTCATATTTCCCTGGATCTCCCTCCTCCCTAGAGACCTCTAATTTATTTGTCCTC	3939
Db	507	GTAAATTCATATTTCTTGTTCTTGTTCTCCTCCTCCCTAGAGACCTCTAATTTATTTGTCCTC	448
Oy	3940	TTCTAGGTTAATTCCTCTTGGATTTGACTTTGTGAGAAGAGGTTGACAGTAGATTAG	3999
Db	447	TTCTAGGTTAATTCCTCTTGGATTTGACTTTGTGAGAAGAGGTTGACAGTAGATTAG	388
Oy	4000	CAAAGTCCCAAGTCCCAAAATTAACAGTGCTGTAGAGTGTCGGGGGAAAAATTAAGTCTAATT	4059
Db	387	CAAAGTCCCAAGTCCCAAAATTAACAGTGCTGTAGAGTGTCGGGGGAAAAATTAAGTCTAATT	328
Oy	4060	TTCCCTCATGAGGATGACACACTGTGAATTCATCTTCAACTGAAGGCCCTGCAGTTCTC	4119
Db	327	TTCCCTCATGAGGATGACACACTGTGAATTCATCTTCAACTGAAGGCCCTGCAGTTCTC	268
Oy	4120	CTAAACCATAGTGTGTTGTTTTCTTTAACAAAGTTTAAGCTAGTGTAAATTAATAA	4179
Db	267	CTAAACCATAGTGTGTTGTTTTCTTTAACAAAGTTTAAGCTAGTGTAAATTAATAA	208
Oy	4180	AAATTTGCTGTCTGCTCACTCCAGCTTGTGTTTAAACCCATTTCAATAGTGTCTG	4239
Db	207	AAATTTGCTGTCTGCTCACTCCAGCTTGTGTTTAAACCCATTTCAATAGTGTCTG	148
Oy	4240	TTGTAAATTCATAACTTTTGGATACCATTTGATGTGTAATAATTTGGTGTCTTGAATAT	4299
Db	147	TTGTAAATTCATAACTTTTGGATACCATTTGATGTGTAATAATTTGGTGTCTTGAATAT	88
Oy	4300	CTTAATAAGAGTCAATTTGTAATTAACATATTTGTGCTGTTAAAAAAAAAAAAAAAAA	4359
Db	87	CTTAATAAGAGTCAATTTGTAATTAACATATTTGTGCTGTTAATTTTGGAGCCCTNTAA	28
Oy	4360	AAAAA 4364	
Db	27	AAAAA 23	

Search completed: March 24, 2005, 16:54:44
Job time : 18305 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 25, 2005, 23:08:29 ; Search time 890 Seconds

(without alignments)
4083.956 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 614
Sequence: 1 MTRSAVFKAAPAGGNPE.....RSRCRLIEQKQTPHRTCRK 614

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 15

Total number of hits satisfying chosen parameters: 17

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US10054935/runat.23032005_144912_26356/app_query.fasta.1.775
-DB=N_Geneseq_16Dec04 -QFMT=fasta -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=100 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=15 -ALIGN=50 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US10054935.@CGN.1.1.470.@runat.23032005_144912_26356 -NCPUB=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database: N_Geneseq_16Dec04:*

1: geneseqn1980s:*\n2: geneseqn1990s:*\n3: geneseqn2000s:*\n4: geneseqn2001as:*\n5: geneseqn2001bs:*\n6: geneseqn2002as:*\n7: geneseqn2002bs:*\n8: geneseqn2003as:*\n9: geneseqn2003bs:*\n10: geneseqn2003cs:*\n11: geneseqn2003ds:*\n12: geneseqn2004as:*\n13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	614	100.0	4372	8	ACC72044 BC10041 g
2	456	74.3	1755	12	Adg86569 Human tum
3	456	74.3	1755	13	Adg84001 Human tum
4	456	74.3	1755	13	Adg83399 Human tum
5	278	45.3	1529	3	AAF18193 Lung canc

6	243	39.6	2887	3	AACT6874	AACT6874 Human ORF
7	188	30.6	1290	11	AD130996	Ad130996 Human CDN
8	155	25.2	3425	12	ADG64204	Adg64204 Novel hum
9	139	22.6	1143	10	ADFS7688	Adf57688 Human pol
10	125	20.4	2887	3	AACT6874	AACT6874 Human ORF
11	118	19.2	385	9	ACH42852	Ach42852 Human foe
12	72	11.7	2106	10	ADFS9738	Adf59738 Human con
13	48	7.8	865	12	ADP84445	Adp84445 Human bre
14	44	7.2	615	6	ABK62818	Abk62818 Rat seque
15	26	4.2	904	6	AA193983	AA193983 Human neu
16	19	3.1	60	6	ABN49572	Abn49572 Human spl
17	17	2.8	903	4	AA194299	AA194299 Human neu

ALIGNMENTS

RESULT 1	
ACC72044	standard; DNA; 4372 BP.
ID	ACC72044
XX	ACC72044;
AC	08-JUL-2003 (first entry)
DT	08-JUL-2003 (first entry)
XX	BCU1041 gene #SEQ ID_65.
DE	Breast cancer; cytostatic; gene therapy; antisense therapy; regulated;
KW	drug discovery; clinical medicine; forensic medicine; gene;
KW	chromosome 7q21.1; ds.
XX	Homo sapiens.
OS	HO2003029421-A2.
PN	10-APR-2003.
XX	02-OCT-2002; 2002MO-US031287.
PD	03-OCT-2001; 2001US-0326526P.
XX	14-MAY-2002; 2002US-00144194.
PR	(ORIG-) ORIGENE TECHNOLOGIES INC.
PA	Sun Z, Li X, Fan W, Kovacs KF, Jay G;
XX	WPI, 2003-381623/36.
DR	P-PSDB; ABR58310.
XX	New isolated human differentially-regulated breast cancer polynucleotide
PT	and polypeptide, useful for diagnosing, staging, prognosticating,
PT	preventing and/or treating diseases and conditions relating to breast
PT	cancer.
XX	Claim 2; SEQ ID NO 65; 127bp + Sequence Listing; English.
XX	The invention relates to isolated polynucleotides which are
XX	differentially-regulated in breast cancer. The methods and compositions
CC	of the present invention are useful for detecting, diagnosing, staging,
CC	monitoring, prognosticating, preventing and/or treating diseases and
CC	conditions relating to breast cancer, and may be used in gene therapy or
CC	antisense therapy. They can also be used in research, drug discovery,
CC	clinical medicine and forensic medicine. Sequences given in records
CC	ACC72012-ACC72074 represent polynucleotides of the invention that are
CC	differentially-regulated in breast cancer. NOTE: The sequence data for
CC	this patent did not form part of the printed specification, but was
CC	obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 4372 BP; 1136 A; 1012 C; 1077 G; 1147 T; 0 U; 0 Other;
SO	Alignment Scores: 0 Length: 4372
Alignment Scores:	0 Length: 4372
Pred. No.:	614.00 Matches: 614
Score:	614.00

Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0
US-10-054-935-2 (1-614) x ACCT2044 (1-4372)			
QY	1	MetThMeArGrSerAlaVal)PheIyAlaAlaAlaProAlaGlyIyAsnProGlu	20
DB	78	ATGACCATATGATGATCCGGGTTCAGAGCGCGCGCCCTCCGCGGCAATCTCTGAG	137
QY	21	GlnArgLeuAspTyrGluArgAlaAlaLeuGlyIyProGluAspGluProGlyAla	40
DB	138	CAGCGATGACTTACAGAGCGGCTGCGGCGCTGAGCGGCGCCCAAGAGAGCTGAGGCG	197
QY	41	AlaGluAlaHisPheLeuProArgHisArgIyAlaLeuGlyIyProGlyProLeuAla	60
DB	198	GCGGAAGCCACTCTCCCGCGGCAACCTTCAAGAGCGCGGCGCCCGCTGCGC	257
QY	61	SerSerGlnGlyIySerProAlaProSerProAlaGlyIyCysGlyIyIyArgIy	80
DB	258	TCCTCCAGAGCGCGGCGCGCTTCCCGCGCGCTGCGCGGCGGCAAGCGCGCGC	317
QY	81	LeuLeuLeuProAlaGlyAlaAlaProGlyIyGlnGluGluSerTrpGlyIySerVal	100
DB	318	TTGTTTCTCCCGCGCGGCGCGCGCGCGGCGGAGAGAGAGCTGAGGCGGTGCGTG	377
QY	101	ProLeuProCysProProProAlaThrIyGlnAlaGlyIyIyGlyIyGluProAla	120
DB	378	CCCTTCCCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	437
QY	121	AlaGlyAlaGlyIyCysSerProArgProIyGlnAlaGlyIyIyGlyIyGluProAla	140
DB	438	GCGGAGC	497
QY	141	SerLeuValAlaAlaIyIyGluProThrProThrAlaGlyAspIyIyGlyIyAla	160
DB	498	TCCTCTGC	557
QY	161	SerProAlaAlaThrAlaSerAspProAlaGlyIyProProProLeuProGlyPro	180
DB	558	TCCCCCGC	617
QY	181	ProProLeuAlaProThrAlaThrAlaGlyIyThrLeuAlaAlaSerGluIyArgTrpIy	200
DB	618	CCACCCCTCGC	677
QY	201	SerMetArgIySerProLeuGlyIyGlyIyGlyIySerGlyIyAlaSerSerGlnAla	220
DB	678	AGTATGAGGAGAGC	737
QY	221	CysLeuIyGlnIyLeuLeuLeuGlnLeuAspLeuIyGluGlnGlnGlnGlnLeu	240
DB	738	TGCTCTCAACAGATCTCTCTGCTGCAATTCGACTCTCATCGAAGAGAGAGAGCAGCTG	797
QY	241	GlnAlaIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIy	260
DB	798	CAGGCGAAGGAGAGAGATCGAGAGCTGAAGTCAAGAGAGAGAGAGAGAGAGAG	857
QY	261	IleGluArgMetGluArgArgMetGlnLeuValIyIyIyIyIyIyIyIyIyIyIyIyIy	280
DB	858	ATTGAACGATGAGAGAGCGGATGAGCTGTAAAGAGAGATTAAGAGAGAGAGAGAG	917
QY	281	IyLeuPheGlnIy	300
DB	918	AAGCTGTTTCAGGAGCTATGAACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG	977
QY	301	IyLeuGlnCysGlnProGluLeuSerGluThrSerGlnThrLeuProProIyProPhe	320
DB	978	AACTGAGAGTGCAGCGGAGCTTTTCCGAGATCCAGAGCTGCTGCTCCAGAGCCCTTC	1037
QY	321	SerCysGlyIyArgSerGlyIy	340

DB	1038	TCATTCGGCGGAGTGCAGAGAGAGACATAAAGAAATCCCATTTGGAGTACAGAAAGA	1097
QY	341	Iy	360
DB	1098	AAGACTCTGTTAAAAAGCTGAGCTCTGAAATTTTCAAAAGTCAAAACAAATCTTAAG	1157
QY	361	HisSerProIleIyGlnGluProCysGlyIySerLeuSerGluThrValCysIyIyIyIyIy	380
DB	1158	CACCTCTTATTAAG	1217
QY	381	LeuArgSerGlnIy	400
DB	1218	TTGAG	1277
QY	401	SerThrProGlnIy	420
DB	1278	TCACACTCCCAAG	1337
QY	421	GluAspLeuProTyrIy	440
DB	1338	GAGATTTGCGGTACCTTTCACACAGAAATGTATTTGTGTGTGTGTGTGTGTGTGTGT	1397
QY	441	ProSerProLeuProLeuArgIy	460
DB	1398	CCATCACCTTACATTAAGGAGATCTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAG	1457
QY	461	LeuMetProSerSerValAlaGlyIy	480
DB	1458	CTGATGCTCAATCAAGTGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1517
QY	481	HisSerValGluProLeuArgAspProAsnProSerAspLeuLeuGluLeuAspAsp	500
DB	1518	CACCTGATGAGAGCTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1577
QY	501	SerValPheSerIy	520
DB	1578	AGTGTGTTTTCAG	1637
QY	521	AspIleGlnArgIy	540
DB	1638	GATATTCAGAGAGTCAAG	1697
QY	541	Iy	560
DB	1698	AAGAAAGAGATTCAGAGATTCGAGCTGAGTTCATTTTCCCTGAGCGAGATGAT	1757
QY	561	ValGluSerLeuMetIleThrProPheLeuProValAlaPheGlyArgProLeuPro	580
DB	1758	GTTGAAAGTTTGAATTAACCCCTTCTGCTGTGTGACATTTGAGACCATTAACCA	1817
QY	581	IyLeuThrProGlnAsnPheGluLeuProTrpLeuAspGluArgSerArgCysArgLeu	600
DB	1818	AAATTAATCCACAGAAATTTTGAAGTACCTGTTGAGTGAAGGCTGAGTCAAGATTG	1877
QY	601	GluIleGlnIy	614
DB	1878	GAGATCCAG	1919
RESULT 2			
ID	AD086569	standard, cDNA, 1755 BP.	
XX	AD086569;		
AC	AD086569;		
DT	07-OCT-2004	(first entry)	
XX			
DE	Human tumour-associated antigenic target (TAT) cDNA sequence #3442.		
XX			
KM	human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;		
XX	cancer; cell proliferative disorder; gene; ss.		
OS	Homo sapiens.		
XX			

Db 1081 CCTATTAAAGAGAACCCCTGTGCTTCTTATCTGAAACTCTTTGTAACTGTAATTGAGG 1140
Qy 383 SerGlnGlnThrProGluYsProArgSerSerValAspThrProProAlaGluSerThr 402
Db 1141 AGCCAGAAAGAACCCCAAGAAAGCCCGGTCTTCAGTGGACACCCACCAAGACTCTCCACT 1200
Qy 403 ProGlnLysGlyProSerThrHisProLysGlnLysAlaIleSerSerGluIleGluAsp 422
Db 1201 CCCCCAAAGAGAGACCCAGACCCATCCCAAGAGAAAGCCCTTCTCAAGTGAATGAAAGAT 1260
Qy 423 LeuProTyrLeuSerThrThrGluMetTyrLeuCyseArgTrpHisGlnProProSer 442
Db 1261 TTGCCCTACTCTTCCACACAGAAATGATTTGTGTGCTTGGACACCGCTCCCCCATCA 1320
Qy 443 ProLeuProLeuArgGluSerSerProLysLysGlnGlnThrValAla 458
Db 1321 CCGTTACCACTTACGGGAATCCTCTCCAAAGAGAGAGAGACTGTAGCA 1368
RESULT 3
ADQ84001
ID ADQ84001 standard; cDNA; 1755 BP.
XX
AC ADQ84001;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human tumour-associated antigenic target (TAT) cDNA sequence #815.
XX
KW human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
KM cancer; cell proliferative disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2004060270-A2.
XX
PD 22-JUL-2004.
XX
PF 15-OCT-2003; 2003WO-US029126.
XX
PR 18-OCT-2002; 2002US-0418988P.
XX
PA (GENTH) GENENTECH INC.
PA (WUTD) WU T D.
PA (ZHOU) ZHOU Y.
PI Wu TD, Zhou Y;
PI
XX
XX
DR MPI, 2004-534300/51.
XX
PT New nucleic acid molecule and encoded polypeptide, for diagnosing,
PT preventing or treating cell proliferative disorders such as cancer.
XX
XX
PS Claim 1; SEQ ID NO 815; 5504bp; English.
XX
XX The present invention describes an isolated tumour-associated antigenic
XX target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide
XX sequences (see SEQ ID NO.1 to 4622); (b) the full-length coding region of
XX (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
XX sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-
XX (c). Also described: (1) an expression vector comprising the above
XX nucleic acid; (2) a host cell comprising the above expression vector; (3)
XX a process for producing a polypeptide; (4) an isolated polypeptide
XX comprising: (a) an amino acid sequence encoded by any of the above
XX nucleotide sequences; (b) an amino acid sequence encoded by the full-
XX length coding region of the above nucleotide sequences; or (c) a sequence
XX having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
XX comprising the above polypeptide fused to a heterologous polypeptide; (6)
XX an isolated antibody that binds to the above polypeptide; (7) a process
XX for producing the antibody; (8) an isolated oligopeptide that binds to
XX the above polypeptide; (9) a tumour-associated antigenic target (TAT)
XX binding organic molecule that binds to the above polypeptide; (10) a
XX composition of matter comprising the above (chimeric) polypeptide,
XX antibody, oligopeptide or TAT binding organic molecule, in combination

CC with a carrier; (11) an article of manufacture comprising a container and
CC the composition of matter contained within the container; (12) methods of
CC inhibiting the growth of a cell that expresses the above protein, where
CC the growth of the cell is at least in part dependent upon a growth
CC potentiating effect of the above protein; (13) a method of
CC therapeutically treating a mammal having a cancerous tumour comprising
CC cells that express the above protein; (14) a method of determining the
CC presence of a protein in a sample suspected of containing the protein
CC described above; (15) methods of diagnosing the presence of a tumour in a
CC mammal; (16) a method for treating or preventing a cell proliferative
CC disorder associated with increased expression or activity of the above
CC protein; and (17) a method of binding an antibody, oligopeptide or
CC organic molecule to a cell that expresses the protein described above.
CC The TAT sequences have cytostatic activities, and can be used in gene
CC therapy. The composition and methods are useful for diagnosing,
CC preventing or treating cancer. The composition is also used for preparing
CC a medicament for the therapeutic treatment or diagnostic detection of a
CC cell proliferative disorder or cancer. The present sequence represents a
CC human TAT cDNA sequence from the present invention.
XX
XX SQ Sequence 1755 BP; 435 A; 502 C; 505 G; 313 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 0 Length: 1755
XX Score: 456.00 Matches: 456
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 74.27% Indels: 0
XX DB: 13 Gaps: 0
XX
XX US-10-054-935-2 (1-614) x ADQ84001 (1-1755)
Qy 3 MetArgSerAlaValPheIysAlaIleAlaIleProAlaGlyIleYsAnProGlnIleArg 22
Db 1 ATGAGATCCGCGGTGTTCAGAGCGCGCGCCCTGCGCGGCAATCCTGACAGCGA 60
Qy 23 LeuAspTyrGlnArgAlaIleAlaIleGlyIleProGlnAspGlnProGlyAlaIleGln 42
Db 61 CTGAGCTACAGAGGCGCTGGCGGCGTGGCGGCGCGAGAGACAGCTGGAGGCGCGAA 120
Qy 43 AlaHisPheLeuProArgHisArgLysLeuLysGlnProGlyProProLeuAlaSerSer 62
Db 121 GCCCACTTCTCCCGGACCGGTAAAGCTCAAGAGACCGGAGCCCGCGCTCTCTCC 180
Qy 63 GlnGlyIleSerProAlaProSerProAlaGlyCyseGlyIleYsGlyIleArgIleLeu 82
Db 181 CAGGGGAGAGCCCGCGGCTTCCCGGCGGCTGGCGGCGGCAAGGCGGCGCTGTGA 240
Qy 83 LeuProAlaGlyAlaIleProGlyIleGlnGlnGlnSerTrpGlyIleSerValProLeu 102
Db 241 CTCCTCCGCGCGGCGCGCCCGCGGAGAGAGAGAGCTGGGCGGTTCGTCCTTG 300
Qy 103 ProCyseProProProAlaThrLysGlnAlaGlyTleGlyIleGlnProAlaIleAlaGly 122
Db 301 CCTGTCCGCCCCCGGACCAAGAGCGGATGGGGGAGCCTGCGAGCGCGA 360
Qy 123 AlaGlyCyseSerProArgProLysTyrGlnAlaValLeuProIleGlnThrGlySerLeu 142
Db 361 GCGGCTGAGCCCGCGGCGGCAAGTACAGCGGTCTGCTTACAGCGGCTCTCTC 420
Qy 143 ValAlaAlaIleAlaIleGlnProThrProTyrAlaGlyAspLysGlyIleAlaIleSerPro 162
Db 421 GTGCGGCGGCGCAAGAGCTTACGCTTGGGTCGGGACAAAGAGTGGGCGGCTCCCC 480
Qy 163 AlaAlaThrAlaSerAspProAlaGlyProProProProLeuProLeuProGlyProProPro 182
Db 481 GTCGCCACCGCTCGACCCGCGGAGACCCCACTACTTGTGCGGCGCGCACCC 540
Qy 183 LeuAlaProThrAlaThrAlaGlyTyrLeuAlaIleSerGlnIleYsArgTyrLysSerMet 202
Db 541 CTGCGCGCACCCCGCACCGCGGAGCCTTGGCGGCGAGCGAGGAGATGGAAGATAG 600
Qy 203 ArgLysSerProLeuGlyIleGlyIleGlyIleSerGlyAlaSerSerGlnAlaIleCyseLeu 222

Score: 278.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 45.28%

Matches: 278
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-054-935-2 (1-614) x AAF18193 (1-1529)

QY 181 ProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGluYArgTrpLys 200
Db 3 CACACCCCTCGGCCACCGCCACCGCCGAGACCTTGGCGGCGACGAGGAGAGAG 62
QY 201 SerMetArgLysSerProLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 220
Db 63 AGTATGAGAGAAAGCCCTCTCTGGGGGTGGTGGCGGCTCGAGGAGCTCAGTCAAGCGCC 122
QY 221 CysLeuLysGlnIleLeuLeuLeuGlnLeuAlaSerGluGluGlnGlnGlnGlnGln 240
Db 123 TGCCTCAAAAGATCTTCTGCTCAATGGACCTCAATCGAAGACGACGACGACGACG 182
QY 241 GlnAlaLysGluLysGluIleGluGluLeuLysSerGluYArgTrpLeuAlaArg 260
Db 183 CAGGCCAAGAGAAAGAGATGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAG 242
QY 261 IleGluArgMetGluYArgMetGlnLeuValLysLysAspAsnGluYArgHis 280
Db 243 ATTGAACGTATGAAAGCGGATCAGCTGTGTAAGAAGATTAAGAGAAAGAGAGAG 302
QY 281 LysLeuPheGlnGlyTrpGluThrGluGluGluGluGluGluGluGluGluGlu 300
Db 303 AAGCTGTTTCAGGGCTATGAAACTGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 362
QY 301 LysLeuGluCysGlnProGluLysSerGluThrSerGlnThrLeuProPhe 320
Db 363 AACTGAGAGTCCAGCGGAGCTTCCGAGACATCCAGACTCTGCTCCCAAGCCCTTC 422
QY 321 SerCysGlyArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArg 340
Db 423 TCATGTGGGGGAGTGGAGAAAGGACATTAAGAAATCCCATTTGGAGATGAGAAAG 482
QY 341 LysThrProValLysLysLeuAlaProGluPheSerLysValLysThrProLys 360
Db 483 AAGACTCTGTATAAAGCTGGCTCTGTAATTTCAAAAGTCAAAACAAAACCTCTAAG 542
QY 361 HisSerProIleLysGluGluProCysGlySerLeuSerGluThrValCysLysArgGlu 380
Db 543 CACTCTCTATTAAGAGAAAGCCGTGGTCTTATCTGAACGTGTTGTAACGTGAA 602
QY 381 LeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProProArgLeu 400
Db 603 TTGAGAGAGCAAGAAAGCCCAAGAAAGCCCGGCTTCTAGTGCACACCCCAAGAGCTC 662
QY 401 SerThrProGluLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIle 420
Db 663 TCCACTCCCCAAAGAGGAGCCAGACCCATCCCAAGAGAGAAAGCTTCTCAAGTGAATA 722
QY 421 GluLysLeuProCysLysSerThrThrGluMetLysLysCysArgTrpHisGlnProPro 440
Db 723 GAAATATTTGCGTACCTTCCACACAGAAATGATTTGTGTGCTGGCACACCTCTCC 782
QY 441 ProSerProLeuArgProLeuArgGluSerSerProLysLysGluGluThrValAla 458
Db 783 CCAATCACCGTTACATTCAGGGAATCTCTCCAAAGAGAGAGAGAGAGAGAGAG 836

RESULT 6
AAC76874
ID AAC76874 standard; cDNA; 2887 BP.
AC AAC76874;
XX 08-FEB-2001 (first entry)
DT
XX Human ORFX ORF2429 polynucleotide sequence SEQ ID NO:4857.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;
KW immunostimulant; thrombotic; coagulant; vasotrophic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antidiabetic;
KW antiviral; antibacterial; antifungal; antineoplastic; antitumor;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.

OS Homo sapiens.

PN W020058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000NO-US008621.

PR 31-MAR-1999; 99US-0127607P.

PR 02-APR-1999; 99US-0127636P.

PR 05-APR-1999; 99US-0127728P.

PR 30-MAR-2000; 2000US-00540763.

PA (CURA-) CURAGEN CORP.

PI Shimkels RA, Leach M;

DR WPI: 2000-602362/57.

DR P-PSDB; AAB42665.

XX Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease.

XX Claim 5; Page 4044-4045; 5507bp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;

CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic;

CC anticonvulsant; antidiabetic; immunosuppressant; immunostimulant;

CC cardiant; thrombotic; coagulant; vasotrophic; antidiabetic; hypotensive;

CC dermatological; immunosuppressive; antineoplastic; antitumor;

CC antiviral; antifungal; antineoplastic; antidiabetic; antianemic. The

CC sequences can be used for determining the presence of or predisposition

CC to, or preventing or treating pathological conditions associated with an

CC ORFX-associated disorder. The nucleic acids can be used to express may be

CC proteins in gene therapy vectors. The proteins and nucleic acids may be

CC used to treat cancers, proliferative disorders, neurodegenerative

CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,

CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester

CC storage, systemic lupus erythematosus, severe combined immunodeficiency

CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune

CC disorders, asthma, aplastic anaemia, burns, wounds, bone and

CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to

CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 0 U; 2 Other;

Alignment Scores:

Score: 6.94e-210

Length: 2887

Matches: 262

Percent Similarity: 99.62%

Best Local Similarity: 99.62%

Query Match: 39.58%

Indels: 1

Mismatches: 1

Indels: 1

Gaps: 0


```
Db      123 GAGAAATTAACCTGAGGTCCAGCCGAGCTTTCCAGACATCCAGACTTGCCTCC 182
Qy      318 LysProPheSerCySGlyAArgSer-GlyLySGlyVHILysArgPheSerProPheGlySe 337
Db      183 AAGCCCTTCTCATGTGGCGGAGTGGAGAAAGGACATMAAGAAATCCCATTTGGAG 242
Qy      337 rThGluArgLysThrProVallySlySLeuAlaProGluPheSerLySVallySThrLy 357
Db      243 TACGAAAGAAAGACTCTCTGTAAAGCTGGCTCTGAATTTTCAAAAGTCAAAACAA 302
Qy      357 sThrProLyHisSerProIleLySGluGluProCyGlySerLeuSerGluThrValCy 377
Db      303 AACTCCCTAACACACTCTCTATTAAAGAGAACCTGTGGTCTTATCTGAACCTGTTG 362
Qy      377 sLySArgGluLeuArgSerGluGluThrProGluLySProArgSerValAspThrPr 397
Db      363 TAAACGTGAATTTGAGGAGCCAGAAACCCAGAAAGCCCGGCTTCAGTGACACCC 422
Qy      397 oProArgLeuSerThrProGluLySGlyProSerThHisProLySGluSAlaPheSe 417
Db      423 ACCAAGACTCTCCACTCCCAAAAGGAGACCCAGACCATCCCAAGAGAAAGCTTCTC 482
Qy      417 rSeiGluILeGluAspLeuProLyLeuSerThrThGluMetTyLeuCySarGTPH 437
Db      483 AAGTGAGATAGAAAGATTGGCGGTACCTTTCACACAGAAATGATTGTGTCGTTGGA 542
Qy      437 sGluProProSerProLeuProLeuArgGluSerSerProLySLeuGluGluThrVa 457
Db      543 CCGAGCCCTCCCCATCCCGTACCATTAACGGAATCCTTCCAAAGAGAGGAGACTGT 602
Qy      457 LAIA 458
Db      603 AGCA 606

RESULT 8
AD064204
ID      AD064204 standard; cDNA; 3425 BP.
XX
AC      AD064204;
DT      07-OCT-2004 (first entry)
XX
DE      Novel human cDNA sequence #1365.
XX
KW      ss: gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
KW      cytoskeletal; gene therapy; diagnostic marker; morbid state; osteoporosis;
KW      neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW      cancer.
XX
OS      Homo sapiens.
XX
PN      EP1440981-A2.
PD      28-JUL-2004.
XX
PF      21-JAN-2004; 2004EP-00001196.
XX
PR      21-JAN-2003; 2003JP-00102206.
XX
PR      09-MAY-2003; 2003JP-00131392.
XX
PA      (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI      Iioagai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI      Yamamoto J, Isono Y, Nagai K, Irie R;
DR      WPI; 2004-535376/52.
XX
DR      P-PDB; AD066392.
XX
PT      Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
XX      Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS      Claim 1; SEQ ID NO 1365; 2449bp; English.
```

```
XX      The invention relates to 2495 novel polynucleotides (I) and their encoded
CC      polypeptides, sequences hybridizing to these nucleotides, sequences
CC      encoding partial polypeptides and sequences having 70% or 90% identity to
CC      the nucleotide and protein sequences. The nucleotides and polypeptides
CC      are useful as diagnostic markers or therapeutic target for the diseases
CC      or morbid states. They are also useful for treating osteoporosis,
CC      neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC      dementia and various cancers. This sequence corresponds to a nucleotide
CC      sequence of the invention.
XX
SQ      Sequence 3425 BP; 922 A; 636 C; 763 G; 1104 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:          5,17e-130          Length:          3425
Score:             155.00              Matches:          155
Percent Similarity: 100.00%            Mismatches:         0
Best Local Similarity: 100.00%          Indels:            0
Query Match:       25.24%              Gaps:              0
DB:                12

US-10-054-935-2 (1-614) x AD064204 (1-3425)

Qy      460 CysLeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrpArg 479
Db      540 TGCTGATGCCATCAAGTGTTCAGAGAGAAACTTCAGTCTTGCGTCTTCTTGAGAG 599
Qy      480 AspHisSerValGluProLeuArgAspProAspProSerSerPheLeuGluAsnLeuAsp 499
Db      600 GACCACTCAATAGAGCCTCTTAAGGAGCCCAAACTCTTCAACCTTTGGAGAACCTGGAT 659
Qy      500 AspSerValPheSerLySArgHisAlaLySLeuGluLeuAspGluLySArgLySArg 519
Db      660 GACAGTGTCTTTTGAAGCGGCATGCAAACTGAGACTGATGAGAGAGAGAGAGAGAG 719
Qy      520 TrpAspILeGluArgIleArgGluGluArgIleLeuGluArgLeuGluLeuArgMetTyr 539
Db      720 TGGGATATTCAAGAGATCAGAGAAACAAAGAAATTTTACAGCGACTGCAGATATGTAT 779
Qy      540 LySLeuLySLeuGlyLeuGluGluSerGluProGluValThrSerPhePheProGluProAsp 559
Db      780 AAAAGAGAAAGAAATTCAGGAATTCGAGCTGAGGTTACCTCATTTTCCCTGAGCCAGAT 839
Qy      560 AspValGluSerLeuMetIleThrProPheLeuProValAlaPheGlyArgProLeu 579
Db      840 GATGTTGAAGGTTGATGATTAACCCCTTCTTGCTGTGTGATGATTGAGACACATTA 899
Qy      580 ProLySLeuThrProGluAsnPheGluLeuProTrpLeuAspGluArgSerArgCySarg 599
Db      900 CCAAAATTACTCCACAGAAATTTGAGCTACCTGTTGATGAGCGTAGCGCATGCAGA 959
Qy      600 LeuGluILeGluLySLeuGluThrProHisArgThrCySargLyS 614
Db      960 TTGAGATCCAGAGAGAGCAAAACCTCACCGGACGTGAGAAA 1004

RESULT 9
ADP57688
ID      ADP57688 standard; cDNA; 1143 BP.
XX
AC      ADP57688;
DT      12-FEB-2004 (first entry)
XX
DE      Human polynucleotide sequence SEQ ID NO:55.
XX
KW      biological activity; genetic engineering; hybridisation probe; oligomer;
KW      primer; chromosome mapping; gene mapping; recombinant protein production;
KW      human; gene; ss.
XX
OS      Homo sapiens.
XX
PN      WO2003080795-A2.
```

```
PD 02-OCT-2003.
XX
XX 09-AUG-2002; 2002MO-US025485.
XX
XX 09-AUG-2001; 2001US-0311261P.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Yang Y, Wang Z, Weng G, Ma Y,
XX
XX WPI; 2003-876918/81.
XX
XX P-PSDB; ADF58688.
XX
XX New polynucleotides, useful as hybridization probes, oligomers or
XX primers, for chromosome or gene mapping, for the recombinant production
XX of proteins, and for generating antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 55; 571bp; English.
XX
XX The present sequence represents an isolated polynucleotide sequence (I)
XX from the present invention, which encodes a polypeptide (II) with
XX biological activity. Also described: (1) a vector comprising (I); (2) an
XX expression vector comprising (I); (3) a host cell genetically engineered
XX to comprise (I) which is operatively associated with a regulatory
XX sequence that modulates expression of (I) in the host cell; (4) a
XX polypeptide (II) encoded by (I); (5) an antibody directed against the
XX polypeptide of (4) and a carrier; (6) an antibody directed against the
XX polypeptide of (4); (7) detecting (I) or the polypeptide of (4) in a
XX sample; (8) identifying a compound that binds to the polypeptide of (4);
XX (9) producing the polypeptide of (4); and (10) a collection of
XX polynucleotides comprising at least one of the polynucleotide sequences
XX (I). The polynucleotides (I) can be used as hybridisation probes,
XX oligomers or primers, for chromosome or gene mapping, for the recombinant
XX production of proteins, and for generating antisense DNA or RNA.
XX
SQ Sequence 1143 BP; 344 A; 287 C; 260 G; 252 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 5.95e-116 Length: 1143
Score: 139.00 Matches: 139
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 22.64% Indels: 0
DB: 10 Gaps: 0
US-10-054-935-2 (1-614) x ADF57688 (1-1143)
QY 476 ProSerTyrArgSerHisSerValGluProLeuArgAspProAsnProSerAspLeuLeu 495
DB |||||
DB 724 CCTCTTGGAGGGACCACTCAGTAGAGCCCTTAAGGAGCAACCAATCCTTGAACCTTTTG 783
QY 496 GluAsnLeuAspAspSerValPheSerTyrArgHisAlaIleuLeuGluLeuAspGluLys 515
DB |||||
DB 784 GAGAACCTGGATACACAGTGTGTTTTCGAACGGCGCATCCAAATCGAGCTGATGAGAAG 843
QY 516 ArgArgLysArgTyrAspIleGlnArgIleArgGluGlnArgIleuGlnArgLeuGln 535
DB |||||
DB 844 AGAAGGAAAAAGATGATATTCAGAGGATCAGGGAACAAAGAAATTTTACAGCAGCTGACG 903
QY 536 LeuArgMetTyrTyrLysLysGlyIleGlnGluSerGluProGluValThrSerPhePhe 555
DB |||||
DB 904 CTCAGATGATGATAAAGAAAGAAAGAAATTCAGGAATCTGAGCTGAGTTACCTCATTTTC 963
QY 556 ProGluProAspAspValGluSerLeuMetIleThrProPheLeuProValAlaIlePhe 575
DB |||||
DB 964 CCGAGACCGAGATATGTGAAGTTTGATGATTAACCCCTTCTTGCTGTGTHAGCATTTT 1023
QY 576 GlyArgProLeuProLysLeuThrProGlnAsnPheGluLeuProTyrLeuAspGluArg 595
DB |||||
DB 1024 GAGCAGACCATTTACCAAAATTAACCTCCACAGAAATTTTGAAGCTACCTGTTGATGAGCGT 1083
QY 596 SerArgCysArgLeuGluIleGlnLysLysGlnThrProHisArgThrCysArgLys 614
|||||
```

```
DB 1084 AGCCGATGCAAGATTGAGATCCAGAAAGCAACACCTCACCAGACGTGTAGAAA 1140
RESULT 10
AAC76874/c
ID AAC76874 standard; cDNA; 2887 BP.
XX
XX AAC76874;
AC
XX
XX 08-FEB-2001 (first entry)
DE
XX
XX Human ORFX ORF2429 polynucleotide sequence SEQ ID NO:4857.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnery; antipariatic; antiparkinsonian; nocrotropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antihumetic; antihypoid;
XX antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antinflammatory disease; coagulation;
XX thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000MO-US008621.
XX
XX 31-MAR-1999; 99US-0127607P.
XX
XX 02-APR-1999; 99US-0127636P.
XX
XX 05-APR-1999; 99US-0127728P.
XX
XX 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M,
XX
XX WPI; 2000-602362/57.
XX
XX P-PSDB; AAB42665.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 5; Page 4044-4045; 5507bp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antipariatic; antiparkinsonian; nocrotropic; neuroprotective; osteopathic;
XX anticonvulsant; antiparkinsonian; nocrotropic; neuroprotective; osteopathic;
XX cardiatic; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX antiviral; antifungal; antihumetic; antihypoid; and antinaemic. The
XX sequences can be used for determining the presence of or predisposition
XX to, or preventing or treating pathological conditions associated with an
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancers, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX disorder, asthma, allergies, aplastic anemia, burns, wounds, bone and
XX cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive
```

XX SQ Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 0 U; 2 Other;
Alignment Scores:
Pred. No.: 7.06e-103 Length: 2887
Score: 125.00 Matches: 125
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.36% Indels: 0
DB: 3 Gaps: 0
US-10-054-935-2 (1-614) x AAC76874 (1-2887)
QY 334 ProPheGlySerThrGluArgGlySerThrProValIysIleuAlaProGluPheSerIys 353
Db 2461 CCATTGGAAGTACGAAAGAAAGACCTCTCTTAAAAAGCTGGCTCTGAAATTTTCAAAA 2402
QY 354 ValIysThrIysThrProIysHisSerProIleIysGluIuProGlySerIeuSer 373
Db 2401 GTCAAAACAAAACCTCCTAAGCACTCTCTTAAAGAGAAACCTGTGTTCTTATCT 2342
QY 374 GluThrValCysIysArgGluIeuArgSerGlnIuThrProGluIysProArgSerSer 393
Db 2341 GAAACTGTTGTAAACGTGAATTAGAGCCCAAGAAACCCCAAGAAAGCCCGGTCTTCA 2282
QY 394 ValAspThrProProArgIeuSerThrProGlnIysGlyProSerThrHisProIysGlu 413
Db 2281 GTGACACACCCACCAAGACTCTCCATCCCAAAAGGAGCCACACCCATCCCAAGAG 2222
QY 414 LysAlaPheSerSerGluIleGluAspIeuProIyIeuSerThrIuMetIyIeu 433
Db 2221 AAAGCTTCTCAAGTGAATGAGATTTGCGTACCTTCCACACAGAAATGTATTTG 2162
QY 434 CysArgTrpHisGlnProProProSerProIeuProIeuArgIuSerSerProIyIys 453
Db 2161 TGTGTTGGACCGACCTCCCTCCATCACCTTACCATTAACGGAATCCTCTCAAGAG 2102
QY 454 GluIuThrValAla 458
Db 2101 GAGGAGACTGTAGCA 2087
RESULT 11
ACH42852
ID ACH42852 standard; cDNA; 365 BP.
XX ACH42852;
XX AC 13-OCT-2003 (first entry)
XX DT 13-OCT-2003 (first entry)
XX DE Human foetal liver/spleen cDNA #68.
XX KW Human; sex; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX OS Homo sapiens.
XX PN US2003073623-A1.
XX PD 17-APR-2003.
XX PF 30-JUL-2001; 2001US-00918995.
XX PR 30-JUL-2001; 2001US-00918995.
XX PA (DRMA/) DRMANAC R T.
XX PA (LABA/) LABAT I.
XX PA (STAC/) STACHE-CRAIN B.
XX PA (DICK/) DICKSON M C.
XX PA (JONE/) JONES L W.
XX PT Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.

XX SQ New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX Claim 1; SEQ ID NO 30064; 44pp; English.
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12769-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20030073623
XX SQ Sequence 365 BP; 122 A; 86 C; 85 G; 72 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.25e-97 Length: 365
Score: 118.00 Matches: 118
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.22% Indels: 0
DB: 9 Gaps: 0
US-10-054-935-2 (1-614) x ACH42852 (1-365)
QY 282 LeuPheGlnGlyTrpGluThrGluArgGluIuThrGluIeuSerGluIyIleIys 301
Db 11 CTGTTACGAGCTTGTGAACCTGAAAGAGAGAGAAACAGAGCTATGTGAAATAATTA 70
QY 302 LeuGluCysGlnProGluIeuSerGluThrSerGlnThrIeuProProIysProPheSer 321
Db 71 CTGAGTGCACAGCCGAGCTTCCGACACATCCAGACTCTGCTCCCAAGCCTTCTCA 130
QY 322 CysGlyArgSerGlyIysGlyHisIysArgIysSerProPheGlySerThrGluArgIys 341
Db 131 TGTGGCGGAGTGGAAAGGACATTAAGAAATCCCATTTGGAAGTACAGAAAGAAAG 190
QY 342 ThrProValIysIysIleuAlaProGluPheSerIysValIysThrIyThrProIyHis 361
Db 191 ACTCCTGTAAAAAGCTGGCTCTCGAATTTTCAAAAGCTCAAAACCTCTTAACAC 250
QY 362 SerProIleIysGluIuProCysGlySerIeuSerGluThrValCysIysArgGluIeu 381
Db 251 TCTCTATTAAAGAGAAACCTGTGTTCTTATCTAATACTGTTTAAACGGAATTG 310
QY 382 ArgSerGlnIuThrProGluIysProArgSerSerValAspThrProArg 399
Db 311 AGGAGCCAAAGAAACCCAGAAAGCCCGGTCTTCAGTGACACCCACCAAGA 364
RESULT 12
ADF59738
ID ADF59738 standard; cDNA; 2106 BP.
XX ADF59738;
XX AC 12-FEB-2004 (first entry)
XX DT 12-FEB-2004 (first entry)
XX DE Human contig polynucleotide sequence SEQ ID NO:2105.
XX biological activity; genetic engineering; hybridisation probe; oligomer;

CC The present invention describes isolated polynucleotide sequences (1),
CC which encode polypeptides (11) with biological activity. Also described:
CC (1) a vector comprising (1), (2) an expression vector comprising (1), (3)
CC a host cell genetically engineered to comprise (1) which is operatively
CC associated with a regulatory sequence that modulates expression of (1) in
CC the host cell; (4) a polypeptide (11) encoded by (1), (5) a composition
CC comprising the polypeptide of (4) and a carrier; (6) an antibody directed
CC against the polypeptide of (4); (7) detecting (1) or the polypeptide of
CC (4) in a sample; (8) identifying a compound that binds to the polypeptide
CC of (4); (9) producing the polypeptide of (4); and (10) a collection of
CC polynucleotides comprising at least one of the polynucleotide sequences
CC (1). The polynucleotides (1) can be used as hybridisation probes,
CC oligomers or primers for chromosome or gene mapping for the recombinant
CC production of proteins, and for generating antisense DNA or RNA. The
CC present sequence represents a human cortig polynucleotide sequence, which
CC is used in an example from the present invention.

US-10-054-935-2 (1-614) X ADF59738 (1-2106)

OY		80	G1VleuenuenueProAlaGlyValAlaAProGlyVGInGInGInuSerTrpGlyVylSer	99
Db		219	GGCTGTATACCCCGGCCGGGGCGCCCCCGGCAGCAGAAGACGTCTGGCGCTTCG	278
OY		100	VAlPLeuLenProCYSPeProPProAlAtmTyrGlnAlaGlytIleGlyGlyLupProAla	116
Db		279	GTGCGCTTGCCGTGTCGCGCCCCCGGCACCAGCAAGCCGGCATTTGGGGGGAGCTGCC	338
OY		120	AlAlaGlyValaGlyCYseSerProAcTProLYtyrGlnAlaValLeuProIleGlnThr	136
Db		339	GCAAGCCGGAGCCGGCTGCAGCCCCCGGCCCAAGTATCAGGCGTGTGCGCCATTCAAGCG	398
OY		140	GISeerLeuValAlaAlaAlaLysGluProThr-ProTprAlaGlyAspLysGlyVal	155
Db		399	GGCTCTCTCGTGGGGGGGCCCAAGAGCCTTAGCCCTGT-GCTGGGGACAAGGGTGGGGGC	457
OY		159	aAla-SerProAlaAlaThrAlaSerAspProAlaGlyProBProLeuProLeuProG	174

RESULT 13	
ADPB4445	
ID	ADPB4445 standard; DNA; 865 BP.
XX	
AC	ADPB4445;
XX	
DT	09-SEP-2004 (first entry)
XX	
DE	Human breast-specific protein coding sequence #44.
XX	
KM	human, breast-specific protein; breast cancer; gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	W02004053077-A2.
XX	
PD	24-JUN-2004.
XX	
PF	05-DEC-2003; 2003WO-US038815.
XX	
PR	05-DEC-2002; 2002US-0431123P.

PA (DIAD-) DIADEXUS INC.
XX
XX
PI Macina RA, Turner LR, Sun Y, Chen H, Rodriguez M;
XX
XX WPI; 2004-468848/44.
DR
DR P-PSDB; ADP84563, ADP84564.
XX
XX
PT New breast specific nucleic acid molecules and polypeptides useful for
PT diagnosing, preventing or treating breast cancer, for producing
PT transgenic animals or cells, or for research purposes.
XX
XX Claim 1; SEQ ID NO 44; 521p; English.

CC The invention comprises the amino acid and coding sequences of human
CC breast-specific proteins. The DNA and protein sequences of the invention
CC are useful for the diagnosis, treatment and prevention of breast cancer.
CC The present DNA sequence encodes a human breast-specific protein of the
CC invention.
XX
XX Sequence 865 BP; 240 A; 180 C; 202 G; 243 T; 0 U; 0 Other;
SQ

Alignment Scores:	
Pred. No.:	1.53e-33
Score:	48.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	7.82%
DB:	12
Length:	865
Matches:	48
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

PA (HISM) HISAMITSU PHARM CO LTD.

PI Nakagawara A;

DR WPI; 2001-565584/63

PT Nucleic acids originating in gene expressed in human neuroblastoma, PT useful as probe or primer in diagnosing prognosis of human neuroblastoma, PT malignancy and susceptibility indicator or tumor marker for anti-cancer PT agents.

PS Claim 1; Page 317; 2979pp; Japanese.

CC The invention relates to novel genes (AI193926-AI197963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes

SQ Sequence 903 BP; 257 A; 169 C; 197 G; 250 T; 0 U; 30 Other;

Alignment Scores:

Pred. No.:	2.06e-05	Length:	903
Score:	17.00	Watch:	30

Score:	17.00	Matches:	30
Percent Similarity:	96.77%	Conservative:	0

Best Local Similarity:	96.77%	Mismatches:	0
Percent Similarity:	96.71%	Conservative:	0

```

Best Local Similarity: 50.77%
Mismatched: 0
Query Match: 2.77%
Indels: 1

```

DB:	4	Gaps:	0
-----	---	-------	---

US-10-054-935-2 (1-614) X AAI94299 (1-903)

QY 585 GlnAsnPhcGluLeuProTyrPleuAspGluArgSerArgCysArgLeu-GluIleGlnIly 604

Db 69 CAGAA TTTGAGCTACCC TGGTGGATGAGCGCTAGCCGATGCAGATTGGGAGATCCAGAA 128

604 blysglnthrprohisargthrcysarglys 614

Db 129 GAAGCAACACCTCACCAGCTGTAGGAA 159

Search completed: March 26, 2005, 03:32:45
Job time : 906 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 26, 2005, 01:18:01 ; Search time 5414 Seconds
(without alignments)
4316.857 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 614
Sequence: 1 MTRSAVFRAAAGAGNPE.....RSRCRLRIKQKTPHRCRK 614

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Word size: 15

Total number of hits satisfying chosen parameters: 406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 100 summaries

Command line parameters:
-MODE=frame_p2n.model -DEV=rlh
-Q/cg2/_USPTO.spool/US10054935/rnuc.23032005_144913_26375/app_query.fasta.1.775
-DB=EST -QFMT=faclap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOCDL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=100
-DOCALLIGN=200 -THR SCORE=quality -THR MIN=15 -ALIGN=50 -MODE=LOCAL -OUTFMT=pic
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054935 @CGC 1.1.3437@rnuc.23032005_144913_26375 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :
EST:
1: gb_est1:
2: gb_est2:
3: gb_est3:
4: gb_est4:
5: gb_est5:
6: gb_est6:
7: gb_est7:
8: gb_est8:
9: gb_est9:
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	332	54.1	3785	3	CR749360 Homo sapi
2	258	42.0	927	5	BU857520 AGENCOURT
3	240	39.1	870	5	BQ229745 AGENCOURT
4	234	38.1	743	7	CN355143 170006001
5	229	37.3	704	7	CN355144 170005326
6	214	34.9	765	5	BQ014717 UT-H-BE1-
7	206	33.6	896	5	BK329236 BK329236
8	203	33.1	891	5	BQ881177 AGENCOURT
9	202	32.9	756	6	CB851322 UI-CF-EC1

10	197	32.1	631	2	BE693636
11	196	31.9	716	6	CD639095
12	194	31.6	593	5	BU781747
13	187	30.5	723	5	BQ187281
14	178	29.0	549	1	A1800794
15	175	28.5	666	6	CA444700
16	175	28.5	721	7	CK300608
17	173	28.2	541	4	BM264268
18	168	27.4	506	7	CR738569
19	167	27.2	583	4	BM263966
20	167	27.2	985	5	BQ711779
21	165	26.9	560	5	EX478274
22	163	26.5	508	5	BQ127983
23	162	26.4	640	6	CA307817
24	160	26.1	644	5	BX952313
25	155	25.2	789	6	CD652041
26	154	25.1	463	1	A1954139
27	153	24.9	460	1	A1582361
28	152	24.8	591	5	BM969790
29	152	24.8	623	4	B1457952
30	150	24.4	451	1	A1831057
31	150	24.4	501	2	AM854130
32	149	24.3	570	7	CN355142
33	148	24.1	841	4	B1517294
34	146	23.8	501	2	BF907636
35	144	23.5	699	1	AV762388
36	142	23.1	428	4	BM728253
37	142	23.1	463	2	AM192854
38	142	23.1	479	2	BF116257
39	139	22.6	551	4	B1667371
40	139	22.6	1032	4	BM475963
41	138	22.5	1087	4	BG421221
42	135	22.0	524	2	BF907542
43	134	21.8	524	2	BF515361
44	133	21.7	580	6	CB268936
45	126	20.5	666	2	AM965141
46	125	20.4	442	1	AM650422
47	125	20.4	1433	3	HSMB0235
48	123	20.0	438	2	BF907618
49	122	19.9	366	1	A1457684
50	122	19.9	377	1	AA635322
51	122	19.9	769	7	CN355147
52	120	19.5	812	7	CK597000
53	120	19.5	900	7	CF111657
54	119	19.4	498	2	BF907618
55	117	19.1	469	1	AA813463
56	116	18.9	426	2	BF888131
57	111	18.1	685	6	CB055193
58	110	17.9	841	7	CO560821
59	109	17.8	385	1	AA316589
60	107	17.4	515	1	BM755716
61	106	17.3	447	2	BF888145
62	106	17.3	809	5	BX098578
63	105	17.1	619	4	BM740068
64	103	16.8	309	7	H11805
65	102	16.4	509	7	H05401
66	101	16.4	531	1	AA402790
67	100	16.3	395	7	H86733
68	100	16.3	397	5	BU150656
69	98	16.0	637	2	BE693616
70	95	15.5	285	1	A1830121
71	95	15.5	287	2	AM726394
72	95	15.5	672	4	AM726394
73	94	15.3	885	7	CK455479
74	93	15.1	399	7	CR773862
75	92	15.0	398	1	A1280516
76	88	14.3	357	1	A1362436
77	86	14.0	1173	5	BQ670290
78	83	13.5	458	6	AA292783
79	83	13.5	508	6	CD288617
80	81	13.2	261	1	AA020921
81	81	13.2	401	1	A1818644
82	75	12.2	450	4	B1289288

BE693636	MR4-BT035
CD639095	AGENCOURT
BU781747	im94e12.y
BQ187281	UI-E-BE1-
A1800794	wg13c04.x
CA444700	UI-H-DH1-
CK300608	UI-E-BE1-
BM264268	1533b05.y
CR738569	CR738569
BM263966	1933b05.x
BQ711779	AGENCOURT
EX478274	DKFZPB686L
BQ127983	1j79g08.y
CA307817	UI-H-FP1-
BX952313	DKFZP781M
CD652041	AGENCOURT
A1954139	wx80b09.x
A1582361	tq67b10.x
BM969790	UI-CF-BU1
B1457952	603198945
A1831057	w162d01.x
AM854130	RC3-CT025
CN355142	170005325
B1517294	603041522
BF907636	QV1-UT009
AV762388	AV762388
BM728253	UI-E-EJ0-
AM192854	x154d12.x
BF116257	7n79h02.x
B1667371	603292145
BM475963	AGENCOURT
BG421221	602451712
BF907542	QV1-UT009
BF515361	UI-H-BE1-
CB268936	1007843.H
AM965141	EST377214
AA650422	ns95h06.s
AM094950	Homo sapi
BM475963	602529340
A1457684	tj48g07.x
AA635322	n67g04.6
CN355147	170004708
CK597000	AGENCOURT
CF111657	Shultzomi
BF907618	QV1-UT009
AA813463	aj34g10.s
BF888131	QV3-TN016
CB055193	NiSC gm08
CO560821	AGENCOURT
AA316589	EST188273
BM755716	K-RST0033
BF888145	QV3-TN016
BX098578	BX098578
BM740068	K-EST0010
H11805	ym11d08.r1
H05401	y180g09.r1
AA402790	z49g11.s
H86733	yt0be10.r1
BU150656	AGENCOURT
BE693616	MR4-BT035
A1830121	w167f04.x
A1362436	qy92h05.x
BQ670290	AGENCOURT
AA292783	z156c11.s
BM726394	UI-E-EJ0-
CK455479	916066.MA
CR773862	DKFZP459K
A1280516	q103e03.x
A1362436	qy92h05.x
BQ670290	AGENCOURT
AA292783	z156c11.s
CD288617	7.A3.Abd
AA020921	z664b09.x
A1818644	wx89b03.x
B1289288	UI-R-DK0-

83	75	12.2	554	2	AM974882	AM974882	EST386987
84	74	12.1	338	7	R99402	Y971F08.r1	
85	74	12.1	682	1	AV745136	AV745136	
86	73	11.9	315	7	H86735	H86735	Yc08e11.r1
87	73	11.9	439	2	BE286441	BE286441	601090980
88	73	11.9	530	2	AM143159	AM143159	EST293455
89	73	11.9	552	6	CB614149	CB614149	AMGNNTC:N
90	73	11.9	601	7	CK38362	CK38362	C629C08-
91	73	11.9	762	7	CK471650	CK471650	AGENCOURT
92	73	11.9	773	5	BU612708	BU612708	UI-M-FRO-
93	73	11.9	3161	3	AK01463	AK01463	Mus muscu
94	72	11.7	287	6	CB710892	CB710892	AMGNNTC:M
95	72	11.7	425	7	R17800	R17800	Y909A07.r1
96	70	11.4	431	2	BE372564	BE372564	601223819
97	70	11.4	515	4	BG143838	BG143838	NCBI43838
98	70	11.4	539	6	CB051590	CB051590	NISC_G101
99	70	11.4	600	6	CA528908	CA528908	8091-58 M
100	70	11.4	686	5	BQ257208	BQ257208	NISC_ko11

ALIGNMENTS

RESULT 1
CR749360
LOCUS CR749360
DEFINITION Homo sapiens mRNA; cDNA DKFZp686J17211 (from clone DKFZp686J17211).
ACCESSION CR749360
VERSION CR749360.1
KEYWORDS GI:51476445
SOURCE HTC.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3785)
Koeber, K., Beyer, A., Mewes, H.W., Weil, B., Aml, C., Osanger, A., Pöb, G., Han, M. and Wiemann, S.
The German CDNA Consortium
COMMENT The German CDNA Consortium
Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Heinrich-Heine-University, Dueseldorf/Germany) within the CDNA sequencing consortium of the German Genome Project. This clone (DKFZp686J17211) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp686J17211
Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/
location/Qualifiers
1..3785
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686J17211"
/tissue_type="rectum tumor"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host DH10B; sites SfiIIA + SfiIIB"
/dev_stage="adult"
/note="hypothetical protein, differentially spliced"
1..3785
/gene="DKFZp686J17211"
107..1348
/gene="DKFZp686J17211"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH18213.1"
/db_xref="GI:51476446"
/translation="MRKSPKGGGGGSSASSAAACIKQILLDILIQQQQIQAAEK EIBELKSRDTLARIEMERRKQIVKDKERHKIKFQGITERERETLSKIKLE CQELSETQTLPPKPSCGSGKHKKRSPFSTERKTVPVKLAPSPSKVKTTPKH

QY 475 lProSerTTPArGAspHisSerValGluProLeuArGAspProAsnProSerAspLeu 495
| | | | |
Db 928 TCCTTCTTGAGGAGGACCACTAGTAGACCTCTTAAGGACCCAAACCTTCAGACCTTTT 987
| | | | |
QY 495 uGluAsnLeuAspAspSerValPheSerLySArGHisAlaLyLeuGluLeuAspGlu 515
| | | | |
Db 988 GGAAGACCTGATACAGTGTGTTTTCGAAGCGGATGCAAAACCTGAGCTGATGAGAA 1047
| | | | |
QY 515 sArGArGLyArGTPArGAspIleGlnArGIlleArGLuGlnArGIlleuGlnArGLeuG 535
| | | | |
Db 1048 GAGAGAGAAAGATGGATATTTCAGAGATTCAGAGAACAAAGATTTCAGCAGCATGCA 1107
| | | | |
QY 535 nLeuArGMeTyrlLySlySlyGlyIleGlnGluSerGluProGluValThrSerPhe 555
| | | | |
Db 1108 GCTCAGATGTATATAAAGAAAGAAATTCAGGATCTGAGCCTGAGGTATACCTATTTT 1167
| | | | |
QY 555 eProGluProAspAspValGluSerLeuMetIleThrProPheLeuProValValAlaP 575
| | | | |
Db 1168 CCTGAGCCAGATGATGTGAAAGTTGATGATTAACCCCTTCCTGCTGTGTAGCAT 1227
| | | | |
QY 575 eGlyArGProLeuProLyLeuThrProGlnAsnPheGluLeuProTrpLeuAspGluAr 595
| | | | |
Db 1228 TGAACGACCATTAACAAATTAATCTCCACAGAAATTTTGAGTACCTGGTTGATGAGCG 1287
| | | | |
QY 595 gSerArGArGLyLeuGluIleGlnLySlyGlnThrProHisArGThrCysArGLyS 614
| | | | |
Db 1288 TAGCCGATGAGATTTGAGATCCAGAAAGAAACCAACCTCAGCGAGCTGTAGGAA 1345
| | | | |
RESULT 2
BUB87520 927 bp mRNA linear EST 16-OCT-2002
LOCUS
DEFINITION
AGENCOURT 10473579 NIH_MGC 107 Homo sapiens cDNA clone
IMAGE:6647374 5', mRNA sequence.
BUB87520
BUB87520.1 GI:24042512
VERSION
BUB87520.1 GI:24042512
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 927)
NIH-MGC http://mgi.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: L1CM2887 row: k column: 22
High quality sequence stop: 667.
FEATURES
source
location/Qualifiers
1..927
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6647374"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_107"
/note="Organ: breast; Vector: pOT57; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

Alignment Scores:
Pred. No.: 2,03e-233 Length: 927
Score: 258.00 Matches: 258
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.02% Indels: 0
Gaps: 0
US-10-054-935-2 (1-614) x BUB87520 (1-927)
QY 271 ValLyLeuAspAspGluLySgluArGHisLyLeuPheGlnGlyTrGluThrGluGlu 290
| | | | |
Db 9 GTAAGAGAGATTAACGAGAAAGAAAGACCAAGCTGTTTTCAGGGCTTGAACAGAGAG 68
| | | | |
QY 291 ArGLyGluThrGluLeuSerGluLySIlleLyLeuGluCysGlnProGluLeuSerGlu 310
| | | | |
Db 69 AGAGAGAAACAGAGCTATCTGAGAAATTAACCTGAGTGCCAGCCGAGCTTTCCGAG 128
| | | | |
QY 311 ThrSerGlnThrLeuProProLyProPheSerCysGlyArGSerGlyLySgluHisLyS 330
| | | | |
Db 129 ACATCCAGACTGCTGCTCCCAAGCCTTCTCATGTGGCGAGTGGAAGGACATTA 188
| | | | |
QY 331 ArGLySerProPheGlySerThrGluArGLySerThrProValLySlyLeuAlaProGlu 350
| | | | |
Db 189 AGGAATCCCATTTGGAGAGTACAGAAAGAAAGACTCTGTAAAAAGCTGCTCTGAA 248
| | | | |
QY 351 PheSerLySValLySThrLySThrProLyHisSerProIleLySgluGluProCysGly 370
| | | | |
Db 249 TTTTCAAAAGTCAAAAGAAAAAAGCTCTTAACACTCTCTTAATAAGGAAACCTGTG 308
| | | | |
QY 371 SerLeuSerGluThrValCysLySArGLyLeuArGSerGlnGluThrProGluLySPro 390
| | | | |
Db 309 TCCTTATCTGAACCTGTTTGTAAACGTGAATTCAGAGCCCAAGAAACCCAGAAACCC 368
| | | | |
QY 391 ArGSerSerValAspThrProProArGLySerThrProGluLySgluProSerThrHis 410
| | | | |
Db 369 CGGCTTCAGTGACACCCCAACCAAGACTCTCCACTCCCAAAAAGGAGACCAAGACCAT 428
| | | | |
QY 411 ProLySgluLySAlaPheSerSerGluIleGluAspLeuProTrpLyLeuSerThrThGlu 430
| | | | |
Db 429 CCAGAGAGAAAGCTTCTCAAGAGATGAAGATTGCCGTTCCTTCCACCAAGAA 488
| | | | |
QY 431 MetTyrlLeuCysArGLTPHisGlnProProProSerProLeuArGLySerSer 450
| | | | |
Db 489 ATGATTTGTGTCTGTGGACACAGCTCCCAATCACCGTTACATTACGGGAATCTCT 548
| | | | |
QY 451 ProLySlySgluGluThrValAlaArGlyLeuMetProSerSerValAlaGlyGluThr 470
| | | | |
Db 549 CCAAGAGAGAGAGACTGATAGCAAGGTGCTGATGCCATCAAGTGTTCAGAGAAACT 608
| | | | |
QY 471 SerValLeuAlaValProSerTTPArGAspHisSerValGluProLeuArGAspProAsn 490
| | | | |
Db 609 TCAGTCTTGCTGTCTCTTCTTGGAGGACACACTAGAGCTCTTAAGGAAACCAAT 668
| | | | |
QY 491 ProSerAspLeuGluGluAsnLeuAspAspSerValPheSerLySArGHisAlaLyLeu 510
| | | | |
Db 669 CTTTCAGACCTTTTGGAGAACCTGATGACAGTGTGTTTCGAAGCGGATGCAAACTG 728
| | | | |
QY 511 GluLeuAspGluLySArGLySArGLTPAspIleGlnArGIlleArGLyGluGln 528
| | | | |
Db 729 GAGCTGATAGAGAGAGAGAAAGATGGGATTTTCAGAGGATCAAGGAAACAA 782
| | | | |
RESULT 3
BQ229745 870 bp mRNA linear EST 02-MAY-2002
LOCUS
DEFINITION
AGENCOURT 7510022 NIH_MGC 92 Homo sapiens cDNA clone IMAGE:6042882
5', mRNA sequence.
BQ229745
BQ229745.1 GI:20411145
VERSION
BQ229745.1 GI:20411145
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 870)
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

FEATURES	Location/Qualifiers
source	1. .870

Alignment Scores:	
Pred. No.:	2.2e-216
Score:	240.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	39.09%
DB:	5
US-10-054-935-2 (1-614) x BQ229745 (1-870)	
	Length: 870
	Matches: 240
	Conservative: 0
	Mismatches: 0
	Indels: 0
	Gaps: 0


```

VERSION      BQ014717.1  GI:19739618
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 765)
AUTHORS      NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs-remail.nih.gov
              Tissue Procurement: Dr. Jose Mercuende
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Clone distribution information can be found
              through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
              Seq. Primer: M13 FORWARD
              POLYA=Yes.

FEATURES
    source
        1..765
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5833443"
            /tissue_type="Chondrosarcoma"
            /dev_stage="Adult"
            /lab_host="DH10B (Life Technologies)"
            /clone_1lb="NCI CGAP ED1"
            /note="Organ: Left Pubic Bone; Vector: pT73-Pac
            (Pharmacia) with a modified polylinker; Site 1: Ecor I,
            Site 2: Not I; NCI CGAP ED1 is a normalized cDNA library
            containing the following tissue(s): Chondrosarcoma cell
            line C55. The library was constructed according to
            Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
            1996. First strand cDNA synthesis was primed with an
            oligo-dT primer containing a Not I site. Double stranded
            cDNA was ligated to an Ecor I adaptor, digested with Not
            I, and cloned directionally into pT73-Pac vector. The
            oligonucleotide used to prime the synthesis of
            first-strand cDNA contains a library tag sequence that is
            located between the Not I site and the (dT)18 tail. The
            sequence tag for this library is GCTCAAGGCT.
            TAG_TISSUE=Chondrosarcoma
            TAG_LIB=UI-H-ED1
            TAG_SEQ=GCTCAAGGCT"

ORIGIN
Alignment Scores:
Pred. No.:      8,54e-192      Length:      765
Score:          214.00         Matches:     251
Percent Similarity: 99.21%     Conservative: 0
Best Local Similarity: 99.21%   Mismatches:  1
Query Match:    34.85%         Indels:      2
DB:             5              Gaps:            0

US-10-054-935-2 (1-614) x BQ014717 (1-765)
QY      291 ArgGluGluThrIleuSerGluValIleLysLeuGluGluGlnProGluIleuSerGlu 310
Db      763 AGAGAGAAACAAGCTATCTGAGAAATTAACCTGAGAGCCAGCCGAGCTTCCGAG 704
QY      311 ThSerGlnThrIleuProProlsProPheSerGlnIleValSerGlyValGly-HisIly 330
Db      703 ACATCCAGACTGCTGCTCCCAAGCCTTCTCATGTGGGCGAGTGAAGG-ACATTA 645
QY      330 sarGlySerProPheGlySerThrGluArgIleThrProValIleLysLeuAlaProGly 350
Db      644 AAGGAATATCCCATTTGAGAGTACAGAAAGAAAGACTCTGTTAAAAAGCTGGCTCTGA 585
QY      350 upheSerIysValIleThrIleThrProIleHisSerProIleLysGluGluProCysGly 370

```

```

Db      584 ATTTTCAAAAGTCMAAACAACAACTCTAGACACTCTTATTAAAGAGAACCTGTGG 525
QY      370 ySerIleuSerGluThrValCylsYargGluLeuArgSerGlnGluThrProGluIysPr 390
Db      524 TTCCTTATCTGAAACCTGTTGTAAAGTGAATTGAGAGACCAAGAAACCCAGAAAGGCC 465
QY      390 oArgSerSerValAspThrProProArgIleuSerThrProGluIysGlyProSerThrH 410
Db      464 CCGGTCTAGTGAAGACACCCACCAAGACTCTCCATCCCAAAAGAGACCCAGACCCA 405
QY      410 sProlsGluIysAlaPheSerSerGluIleGluAspLeuProTyLeuSerThrThrG 430
Db      404 TCCCAAGGAGAAAGCCTTCTCAAGTGAAGTAAAGATTGGCGTACCTTTCCACCAAG 345
QY      430 uMetTyLeuCySarGlyTrpHisGlnProProSerProLeuProLeuArgIleuSer 450
Db      344 AATGATTTGTGTGTCGTGGACACAGCCTCCCATCAGCTTACATTACGGGAATCTTC 285
QY      450 rProlsIysGluGluThrValAlaArgCylsMetProSerSerValAlaGlyGluTh 470
Db      284 TCCAAAGAAAGAGAGACTGTAGCAAGGTCTGATGCATCAAGTTCAGAGAGAAAC 225
QY      470 rSerValIleAlaValProSerTrpArgAspHisSerValGluProLeuArgAspProAs 490
Db      224 TTCAGTCTTGCTGCTGTTCTTCTTGGAGGACCACTAGTAGACCTCTTAAGGACCCAA 165
QY      490 nProSerAspLeuLeuGluAsnLeuAspSerValPheSerIysArgHisAlaIysIle 510
Db      164 TCTTTCAGACCTTTTGGAGAACCTGTGATGACAGTGTGTTTGCAGACGCGATCAAAACT 105
QY      510 uGluLeuAspGluValArgArgIleArgTrpAspIleGlnArgIleArgGluGlnArg 530
Db      104 GAGGCTGAGTGAAGAAAGAAAGAAAGATTCAGAGGATCAGGAAACAAAGAAAT 45
QY      530 eLeuGlnArgLeuGlnLeuArgMetTyIysIysIys 542
Db      44 TTTTACGACGACTGCACTCAGATTGATTAATAAAAAA 8

RESULT 7
BX329236/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Genoscope - Centre National de Sequencage
2 rue Gaston Creneau, CP 5706 - 91057 Evry cedex - FRANCE
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and Ecor V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
8989.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?b=CS0BA1036ZB04_CS03386_1&c=8989.f

FEATURES
    source
        1..896
            /organism="Homo sapiens"
            /mol_type="mRNA"

```

```

/db xref="taxon:9606"
/clone="CSODK002YN11"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELa CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

Alignment Scores:	
Pred. No.:	3,81e-184
Score:	205.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	33.55%
DB:	5
Length:	836
Matches:	206
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-054-935-2 (1-614) x BX329236 (1-896

QY	163	AlaAlaThrAlaAspAspProAlaGlyProProProLeuProLeuProGlyProProPro	182
Db	640	GCTGCCACCGCCTCGGACCCGGGCGGACCCCACTCACTCTGCCCCGGCCCAACC	581
QY	183	LeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGlyArgTyrGlySerMet	2020
Db	580	CTCGCGCCCAACCGCAACCCCGGAGCCCTCGCGCGCCAGCGAGGCGAGATGGAAGAGTATG	521
QY	203	ArgLysSerProLeuGlyGlyGlyGlySerGlyAlaSerSerGlnAlaAlaCysLeu	2222
Db	520	AGGAAGAGCCCTCTCGGGGGGTGTGGCGGCTCGGGAGCCTCCAGTCAAGCGCGCTGCTC	461
QY	223	LysGlnIleLeuLeuLeuGlnLeuAspLeuIleGluGlnGlnGlnGlnLeuGlnAla	2424
Db	460	AAACAGATCTCTCTGCTGCAATTGGACCTCATGGAACAGCAGCAGCAGCTGCAGGCC	401
QY	243	LysGluLysGluIleGluGluLeuLeuLysSerGluArgAspThrLeuLeuAlaArgIleGlu	2626
Db	400	AAGGAAAAAGAGATCGAGAGCTGAACTCAGAGAGAGACGCTCTCTGCTCGGATTGAA	341
QY	263	ArgMetGluArgArgMetGlnLeuValLysLysAspAsnGluLysGluArgHisLysLeu	2828
Db	340	CGTATGGAAAGCCGGATGCAGCTGGTAAAGAGAGATACAGAAAGAAAGGCAACAGCTG	281
QY	283	PheGlnGlyTyrGluThrGluGluArgGluGluThrGluLeuSerGluLysIleLysLeu	3020
Db	280	TTTCGAGGCTATGAAACTGAAGAGAGAGAAACAGAGCTATCTGGAATAATTAACTG	221
QY	303	GluCysGlnProGluLeuSerGluThrSerGlnThrLeuProProLysPhePheSerCys	3222
Db	220	GAGTCCACGCGGAGCTTTCGAGAGATCCCAACTCTGCTCCCAAGCCCTTCTCATGT	161
QY	323	GlyAspSerGlyLysGlyHisLysAspGlySerProPheGlySerThrGluArgLysThr	3424
Db	160	GGGCGGAGTGGAAAGGACATTAAGCAATCCCACTTGGAGTACAGAAAGAAAGACT	101
QY	343	ProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysHisSer	3626
Db	100	CCTGTTAAAAAACTGGCTCCTGCAATTTTCAAAAGTCAAAACAAAACTCTTAAGACTCT	41
QY	363	ProIleLysGluGluPro	368
Db	40	CTATTAAAGAGAAACC	23

RESULT	8
BO881177	
LOCUS	
DEFINITION	BO881177 891 bp mRNA linear EST 16-AUG-2002
ACCESSION	AGNCOURT_8065588 NIH_MGC_102 Homo sapiens CDNA clone IMAGE:6212597
VERSION	5', mRNA sequence.
KEYWORDS	BO881177 GI:22273185

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens (human)	Homo sapiens	Eukaryotes: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
1 (bases 1 to 891)		NIH-MGC http://mgc.nci.nih.gov/ .				
Unpublished (1999)		National Institutes of Health, Mammalian Gene Collection (MGC)				
Contact: Robert Strausberg, Ph.D.						
Email: cgabbs-remail.nih.gov						
Tissue Procurement: ATCC						
CDNA Library Preparation: Rubin Laboratory						
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)						
DNA Sequencing by: Agencourt Bioscience Corporation						
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:						
http://image.llnl.gov						
Plate: L1CM2376	row: p	column: 06				
High quality sequence	stop: 601.					
Location/Qualifiers						
1..891						

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6212597"
/lisuse_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH MGC 102"
/note="Organ: salivary gland; Vector: pOMB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NH MGC library."

```

Alignment Scores:		
Pred. No.:	2.64e-181	Length: 891
Score:	203.00	Matches: 203
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	33.06%	Indels: 0
DB:	5	Gaps: 0

QY	256	Thrlleuenu1aaqgll1egiuarGMeGluabrgprmetG1neuVal1yrlvSaAspn	275
Db	5	ACGCTCCTGTCGATATTGAACGTATGAAAGGCGGATGACGCTGTTAAAGAAATTAAC	64
QY	276	Glu1yrgsluarGfH1sly1sLeuPheG1ng1TYrG1uthrg1uglu1AargGlu1uthrg1u	295
Db	65	GAGAAAGAAAGGCAAGCTGTTTCAGGGCTATATAACTGAAGAGAGAGAAACAGAG	124
QY	296	LeuSerg1uyl1ely1sLeuGlu1CyGg1nProGlu1euSerg1uthrSerg1uthrleu	315
Db	125	CTATCTGAGAAATTAATTAACGTGAAGTGCAGCGGAGCTTCCGAGACATCCAGACTCTG	184
QY	316	ProProly1sProPheSerCyG1yAargSerg1yl1y1sG1y1hly1sVal1y1sSerProPhe	335
Db	185	CCTCCCAAGCCCTTCTCATGTGGCGGAGTGGAAAGGACATTAAGAAGAAATCCCCATTT	244
QY	336	GlySerThrGluAargly1uthrProVal1y1sleu1a1aProGlu1uPheSer1yVal1y1s	355
Db	245	GGAAGTACAGAAAGAAAGACTCCTGTTAAAGCTGGCTCCTGAATTTTCAAAGTCAAA	304
QY	356	Thrl1y1sThrProly1sH1s1sSerPro11ely1sGlu1uProCySg1ySer1euSerg1uthr	375
Db	305	ACAAAACCTCTAAGCACTCTCTCATTTAAAGAGAAACCTGTGTCTCTTAATCGAAACT	364
QY	376	ValCy1y1sVal1y1sGlu1euAargSerg1ng1uthrProGlu1y1sProAargSerVal1asp	395

```
Db 365 GTTTGTAACGTGAATTGAGACCCAGAAACCCGAGAAAGCCCGGTTCATGAGGAC 424
Qy 396 ThrProArgLeuSerThrProGlnLySGlyProSerThrHisProLySGlnLySAla 415
Db 425 ACCCCACCAAGACTCTCCACTCCCAAAAAGGAGCCAGACCCATCCCAAGAGAAAGCC 484
Qy 416 PheSerSerGlnIleGlnAspLeuProTyrLeuSerThrThrglnMetTyrLeuCyArg 435
Db 485 TTCTCAAGTAGAGATTAAGATTTGCGGTTCCTTCCACCAAGAAATGATTTGTGTCGT 544
Qy 436 TTPHISGLIPROPROSERPROLEUARGLEUARGLEUSERPGLYSGLYGLU 455
Db 545 TGGCACCAAGCTTCCCATCATCCGTTACCATTTACGGGAATCCTTCCAAAGAGAGAG 604
Qy 456 ThrValAla 458
Db 605 ACTGTAGCA 613
```

```
RESULT 9
CB851322 756 bp mRNA linear EST 22-APR-2003
LOCUS UI-CF-EC1-ady-c-18-0-UI .s1 UI-CF-EC1 Homo sapiens cDNA clone
DEFINITION UI-CF-EC1-ady-c-18-0-UI 3', mRNA sequence.
ACCESSION CB851322
VERSION CB851322.1 GI:30046094
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 756)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.regen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=No.
```

FEATURES

source

```
1..756
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EC1-ady-c-18-0-UI"
/tissue_type="Lung"
/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_1ib="UI-CF-EC1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: Ecor I; Site_2: Not I;
UI-CF-EC1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an Ecor I
```

adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AAGTGCTTAC.

TAG_SEQ=None found"

```
ORIGIN
Alignment Scores:
Pred. No.: 2e-180 Length: 756
Score: 202.00 Matches: 202
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 32.90% Indels: 0
DB: Gaps: 0
US-10-054-935-2 (1-614) x CB851322 (1-756)
```

```
Qy 236 GlnGlnGlnGlnLeuGlnAlaLySGlnLyGlnIleGlnGlnLeuLySerGlnArgAsp 255
Db 8 CAGCAGCAGCAGCTGCAGGCCAAGAAAGAGATCAGAGAGCTGAAGTCAGAGAGAGAC 67
Qy 256 ThrLeuLeuAlaArgIleGlnArgMetGlnArgArgMetGlnLeuValLySlyAspAsn 275
Db 68 ACGCTCTTGCTCGAGTGAAGATGAGAAAGCGAGTCAGCTGTGAAGAGATTAAC 127
Qy 276 GlnLySGlnArgHisLySLeuPheGlnLyTyrGlnThrglnGlnArgGlnGlnGln 295
Db 128 GAGAAAGAAAGGACACAGAGCTTTTCAGGGCTAAGAAATCGAAGAGAGAGAAACAGAG 187
Qy 296 LeuSerGlnLySleLySLeuGlnCyGlnProGlnLeuSerGlnThrSerGlnThrLeu 315
Db 188 CTATCTGAGAAATTAAGTGAAGTCCAGGCCGAGCTTTCAGACATCCAGACTCTG 247
Qy 316 ProProLySProPheSerCysGlnArgSerGlnLySArgLySlyArgLySArgProPhe 335
Db 248 CTTCCAGAGCCCTTCATGATGAGCGAGTGGAAAGGACATTAAGAGAAATCCCATTT 307
Qy 336 GlySerThrGlnArgLySThrProValLySlySLeuAlaProGlnPheSerLySValys 355
Db 308 GGAAGTACGAAAGAAAGAAAGCTCCTGTTAAAGAAAGCTGCTCTGAATTTCAAAAGTCAA 367
Qy 356 ThrLySThrProLySHisSerProLleLySLeuGlnProCysGlnSerLeuSerGlnThr 375
Db 368 ACAAAAAGCTCTTAAGACCTCTCTATTAAAGAGAAAGCCGTGTGTTCTTATCTGAAC 427
Qy 376 ValCyluAlaArgGlnLeuArgSerGlnGlnThrProGlnLySProArgSerSerValAsp 395
Db 428 GTTTGTAACGTGAATTGAGAGCCAGAAACCCAGAAAGCCCGGTCTTCAAGTGAAC 487
Qy 396 ThrProArgLeuSerThrProGlnLySGlyProSerThrHisProLySGlnLySAla 415
Db 488 ACCCCACCAAGACTCTCCACTCCCAAAAAGGAGCCAGACCCATCCCAAGAGAAAGCC 547
Qy 416 PheSerSerGlnIleGlnAspLeuProTyrLeuSerThrThrglnMetTyrLeuCyArg 435
Db 548 TTCTCAAGTAGATTAAGATTTGCGGTACCTTCCACCAAGAAATGATTTGTGTCGT 607
Qy 436 TTPHIS 437
Db 608 TGGCAC 613
```

RESULT 10

LOCUS BE693636 631 bp mRNA linear EST 11-SEP-2000

DEFINITION MK4-BT0358-120600-009-c01 BT0358 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE693636

VERSION BE693636.1 GI:10080796

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	ALTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
1	(bases 1 to 631)	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
2	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunstein, A. J., deOliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	20202663	10737800	
3	Contact: Simpson A.J.G.	Laboratory of Cancer Genetics	Ludwig Institute for Cancer Research	Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil		
4	Tel: +55-11-2704922					
5	Fax: +55-11-2707001	Email: asimpson@ludwig.org.br	This sequence was derived from the FAPESP/ICR Human Cancer Genome Project. This entry can be seen in the following URL	(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-MR4-BT0358-1200-009-001&t3=2000-06-12&t4=1)		
6	Seq primer: puc 18 forward	High quality sequence start: 17	High quality sequence stop: 606.			
7	Location/Qualifiers					
8	1. .631					
9	/organism="Homo sapiens"					
10	/mol_type="mRNA"					
11	/db_xref="taxon:9606"					
12	/dev_stage="Adult"					
13	/clone_lib="BT0358"					
14	/note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."					
15	ALIGNMENT SCORES:					
16	Prod. No.:	9.19e-176	Length:	631		
17	Score:	197.00	Matches:	197		
18	Percent Similarity:	100.00%	Conservative:	0		
19	Best Local Similarity:	100.00%	Mismatches:	0		
20	Query Match:	32.08%	Indels:	0		
21		2	Gaps:	0		
22	US-10-054-935-2 (1-614) x BE693636 (1-631)					
23	262	GLUARGWETGLUARGARMEGLNLEUVALYSLYSAPAENGILUYGLUARGHISLYS	281			
24	23	GAAAGTATGCAAAAGCGGATGACACGTGTAAGAAGATACGAAAGAAAGACACAG	82			
25	282	LEUHEGNGLYYRYGLUUTHRGUUGLUARGGLUUGLUTHRGUULENSEGLUYELI	301			
26	83	CTGTTCACGGGCTVATGAAATCGAAGAGAGAGAGAAACAGGCTATCGAATAATTAA	142			
27	302	LEUGLUCYSGINPRTGULLEUSEGLUTHRSERGINTHLEAPROFOLYPROPHESER	321			
28	143	CTGAGATGCCAGCCGAGCTTTCGAGACATCCAGACTCTGCTCCAAACCTTCTCA	202			
29	322	CYSGLYARGSERGLYLYGLYHISLYEARGLYSERPROHEGLYSETTRHGLUARYLS	341			
30	203	TGTGGCGCGATGGAAAGGACATPAAAGAAATCCCATTTGGAAGTACAGAAAGAAAG	262			
31	342	THRPVALYSLYSLLEUVALAPROGLUPHESERLYSLVSTHRYLSTHPROLYHIS	361			
32	263	ACTCTGTAAAGAGCTGGCTCTCGAATTTTCAAAAGTCAAAACAAAATCTCTTAGAC	322			

Qy	362	SeqProIleysgInguPProCysGlySetLeusSerJuthrValCylsYsArgJlueu	381
Db	323	TCTCCATTAAAGAGAACCTGTGGTTCCTTATCTGAACCTGTTTATAACGTGAATTG	382
Qy	382	ArgSerGlnJuthrProGluYsProArgSerServAlasPhThrProProArgLeuSer	401
Db	383	AGAGGCCAAGAAACCAGAAAAAGCCCCGGGTCTTCAGTGAGACCCCAACAAGACTCTCC	442
Qy	402	ThrProGlnYsGlyProSerThrHisProLyGlnUlysAlaPheSerSerGlnJlueGln	421
Db	443	ACTCCCAAAGGAGACCGACACCCCATCCCAAGAGAAACCTTCTCAAGTGAAGTAGAA	502
Qy	422	AspLeuProTyrlseuSerThrThrGluMetTyrlseuCysArgTrpHisGlnProPro	441
Db	503	GATTGCCGTAACCTTTCCACCACAGAATATATTGTGTGCTTGACACAGCTCCCCCA	562
Qy	442	SerProIleuProIleuArgJlueSerSerProLyIsGlnJuthrValAla	458
Db	563	TCAACGGTATCACATTACGGGAATCCTCTCCAAAGAGAGAGACTGTAGCA	613
RESULT	11		
LOCUS	CD639095	716 bp	mRNA linear EST 17-JUN-2003
DEFINITION	ABENCOUR 14532950 NIH MGC 191 Homo sapiens CDNA clone		
VERSION	CD639095	IMAGE:30418364 5', mRNA sequence.	
KEYWORDS	CD639095.1	GI:31805567	
SOURCE ORGANISM	EST. Homo sapiens (human)		
REFERENCE AUTHORS	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL TITLE	NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgaabs@mai.nih.gov Tissue Procurement: Narayan Bhat CDNA Library Preparation: CLONTECH Laboratories, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: NDCM221 row: d column: 21 High quality sequence stop: 569.		
FEATURES			
SOURCE			
	1..716	/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:30418364"	
		/ribose_type="Pooled"	
		/lab_host="DH10B (TI phage-resistant)"	
		/clone_lib="NIH_MGC_191"	
		/note="Vector: pDNR-LIB; Site 1: SfiI (ggccatcgcc); Site 2: SfiI (ggcgctcgccc); Library is oligo-dt primed and directionally cloned. PBMC - Peripheral Blood Mononuclear Cells. RNA was pooled from 3/6hour stimulation with PMA adn Ionomycin. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGAGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCCGACATG-dr(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.69 kb (range 0.70-5.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."	

ORIGIN

Alignment Scores:

Pred. No.: 9.22e-175 Length: 716

Score: 196.00 Matches: 196

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 31.92% Indels: 0

DB: 6 Gaps: 0

US-10-054-935-2 (1-614) x CD639095 (1-716)

QY 410 HisProlysgluValaPheSerSergluIleGluAspLeuProTyrLeuSertThr 429

DB 3 CATCCAGAGAGAAAGCCTTCTCAAGTAGATGAAATTTGCCGTACCTTCCACACACA 62

QY 430 GluMetTyrLeuCyAspGTPHISGInProProSerProleuProleuArgGluSer 449

DB 63 GAAATGATTTTGTCGTCGTGGACACAGCCTCCCATCACCCTTACCTTACGGGAATCC 122

QY 450 SerProlysgluValaValaValaValaValaValaValaValaValaValaValaVala 469

DB 123 TCTCCAAAG 182

QY 470 ThrSerValleuValaValaProSertTyrAspHisSerValaGluProleuValaAspPro 489

DB 183 ACTTCAGTCTTGCTGCT 242

QY 490 AspProSerAspLeuGluValaValaValaValaValaValaValaValaValaValaVala 509

DB 243 AATCCTTTCAGACCTTTTGGAGAACCTGGATGACAGCTGCTTTTGGAGACCGCATGCAAAA 302

QY 510 LeuGluLeuAspGluValaValaValaValaValaValaValaValaValaValaVala 529

DB 303 CTGAGCTGATGAG 362

QY 530 IleLeuGluArgLeuGluLeuValaValaValaValaValaValaValaValaValaVala 549

DB 363 ATTTTACAG 422

QY 550 GluValThrSerPhePheProGluProAspAspValaGluSerLeuMetIleThrProPhe 569

DB 423 GAGGTTACCTCATTTTCT 482

QY 570 LeuProValaValaPheGluArgProleuProleuProleuProleuProleuProleu 589

DB 483 TTGCTGTTTACATTTTGGACGACCATTAACAATACTCCACAGATTTTGGAGCTA 542

QY 590 ProTyrLeuAspGluValaValaValaValaValaValaValaValaValaValaVala 605

DB 543 CCTGCTTGATGAG 590

RESULT 12

LOCUS BU781747 593 bp mRNA EST 11-OCT-2002

DEFINITION Im94e12.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:6122614

ACCESSION BU781747

VERSION BU781747.1 GI:23824324

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Melton, P., Brown, J., Keny, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Scaerle, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Bliststein, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T., Jackson, Y., and Bowers, Y.

TITLE Endocrine Pancreas Consortium

JOURNAL Unpublished (2000)

COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biochem.harvard.edu

Library was constructed by Dr. J. Ferrer in vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

Seq primer: -40RP from Gibco

High quality sequence stop: 430.

Location/Qualifiers

1. 593

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6122614"

/issue_type="Insulinoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="Human insulinoma"

/note="Organ: pancreas; Vector: pBluescript SK-; Site:1: XhoI, Site:2: EcoRI; Constructed with lambda ZapII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol

(http://genome.wustl.edu/est/lambda_protocol.shtml).

Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

ORIGIN

Alignment Scores:

Pred. No.: 6.04e-173 Length: 593

Score: 194.00 Matches: 194

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 31.60% Indels: 0

DB: 5 Gaps: 0

US-10-054-935-2 (1-614) x BU781747 (1-593)

QY 187 AlaThrAlaGlyThrLeuValaValaValaValaValaValaValaValaValaValaVala 206

DB 10 GCCACCGCGGAGCCTTGGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69

QY 207 LeuGluValaGlyGlySerGluValaSerSergluAlaValaValaValaValaVala 226

DB 70 CTCGGGGGTGGTGGCGGCTCGGAGGCTCAGTCAAGCGCGCTGCTCAAGAGATCCTT 129

QY 227 LeuLeuGluLeuAspLeuIleGluGluGluGluGluGluGluGluGluGluGluGlu 246

DB 130 CTGCTCAATTTGAGCTTCAATCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 189

QY 247 IleGluGluLeuLeuSerGluValaAspThrLeuValaValaValaValaValaVala 266

DB 190 ATCAGAGAGCTGAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 249

QY 267 ArgMetGluLeuValaValaValaValaValaValaValaValaValaValaValaVala 286

DB 250 CGGATGACAGCTGTTAAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 309

QY 287 GluThrGluGluValaValaValaValaValaValaValaValaValaValaValaVala 306

DB 310 GAAACTGAG 369

QY 307 GluLeuSerGluThrSerGluThrLeuProProProProProProProProProProPro 326

DB 370 GAGCTTTCGAGACATCCAGACTGCTGCTCCAGACCTTTCATGATGGGCGAGATGA 429

QY 327 LygGlyHisLysArgLysSerProPheGlySerThrGluArgLysThrProValLysLys 346
| | | | |
Db 430 AAGGACATTAAGAAATCCCATTTGGAAGTACAGAAAGAAAGATCTCTGTTAAAAAG 489
| | | | |
QY 347 LeuAlaProGluProHisSerLysValLysThrLysThrProLysHisSerProLysGlu 366
| | | | |
Db 490 CTGGCTCTCTAATTTTAAAAAGTCAAAACAAACCTCTTAAGCACTCTCTATTAAAGAG 549
| | | | |
QY 367 GluProCysGlySerLeuSerGluThrValCysLysArgLys 380
| | | | |
Db 550 GAACCTGTGTCTTCTTCTGAACCTGTTTGTAAAGCTGAA 591
| | | | |
RESULT 13
BO187281 723 bp mRNA linear EST 30-APR-2002
LOCUS BO187281
DEFINITION UT-E-BJ1-aj2-b-15-0-UT, r1 UT-E-BJ1 Homo sapiens cDNA clone
| | | | |
ACCESSION BO187281
| | | | |
VERSION BO187281.1 GI:20362832
| | | | |
KEYWORDS EST.
| | | | |
SOURCE Homo sapiens (human)
| | | | |
ORGANISM Homo sapiens
| | | | |
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
| | | | |
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
| | | | |
TITLE 1 (bases 1 to 723)
| | | | |
BONALDO, M.F., LEMMON, G. and SOARES, M.B.
| | | | |
Normalization and subtraction: two approaches to facilitate gene
| | | | |
discovery
| | | | |
Genome Res. 6 (9), 791-806 (1996)
| | | | |
JOURNAL 97044477
| | | | |
MEDLINE 8889548
| | | | |
COMMENT Coordinated Laboratory for Computational Genomics
| | | | |
Contact: Soares, MB
| | | | |
University of Iowa
| | | | |
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
| | | | |
TEL: 319 335 8250
| | | | |
Fax: 319 335 9565
| | | | |
Email: bento-soares@uiowa.edu
| | | | |
Tissue Procurement: Dr. Gregg Hageman
| | | | |
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
| | | | |
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
| | | | |
Clone Distribution: Researchers may obtain clones from Research
| | | | |
Genetics (www.resgen.com).
| | | | |
Seq primer: M13 REVERSE.
| | | | |
FEATURES
| | | | |
Source Location/Qualifiers
| | | | |
1..723
| | | | |
/organism="Homo sapiens"
| | | | |
/mol_type="mRNA"
| | | | |
/db_xref="taxon:9606"
| | | | |
/clone="UT-E-BJ1-aj2-b-15-0-UT"
| | | | |
/tissue_type="fetal eyes, lens, eye anterior segment,
| | | | |
optic nerve, retina, Retina Foveal and Macular, RPE and
| | | | |
Choroid"
| | | | |
/dev_stage="fetal and adult"
| | | | |
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
| | | | |
/clone_lib="UT-E-BJ1"
| | | | |
/note="Organ: eye; Vector: pTTT3-Pac (Pharmacia) with a
| | | | |
modified polylinker; Site 1: Bcor I; Site 2: Not I;
| | | | |
UT-E-BJ1 is a subtracted cDNA library constructed
| | | | |
according to Bonaldo, Lemmon and Soares, Genome Research,
| | | | |
6:791-806, 1996. First strand cDNA synthesis was primed
| | | | |
with an oligo-dT primer containing a Not I site. Double
| | | | |
stranded cDNA was ligated to an Bcor I adaptor, digested
| | | | |
with Not I, and cloned directionally into pTTT3-Pac
| | | | |
vector. The oligonucleotide used to prime the synthesis of
| | | | |
first-strand cDNA contains a library tag sequence that is
| | | | |
located between the Not I site and the (dT)18 tail. The
| | | | |
sequence tags for this library are: fetal eyes,
| | | | |
AGATCAAGA, lens, CGATTAGCA, eye anterior segment,
| | | | |
AATCCGCAAT, optic nerve, CCATTAGTG, retina, CCGCG, Retina
| | | | |
Foveal and Macular, GTCC, RPE and Choroid, ACCTA. This

ORIGIN
| | | | |
Alignment Scores:
| | | | |
Pred. No.: 3,16e-166 Length: 723
| | | | |
Score: 187.00 Matches: 202
| | | | |
Percent Similarity: 99.51% Conservative: 0
| | | | |
Best Local Similarity: 99.51% Mismatches: 1
| | | | |
Query Match: 30.46% Indels: 1
| | | | |
DB: 5 Gaps: 0
| | | | |
US-10-054-935-2 (1-614) x BO187281 (1-723)
| | | | |
QY 249 GluLeuLysSerGluArgAspThrLeuAlaArgIleGluArgMetGluArgArgMet 268
| | | | |
Db 2 GAGCTGAAGTCAGAGAGAGACAGCTCTTCTCGATTAAAGCTAT-GAAAGCCGATG 60
| | | | |
QY 269 GluLeuValLysLysAspAsnGluLysGluArgHisLysLeuPheGlnGlyTyrGluThr 288
| | | | |
Db 61 CAGCTGTAAAGAGAGATTAAGAGAAAGAGCAAGCTGTTTCAGGCTATGAAGACT 120
| | | | |
QY 289 GluGluArgGluGluThrGluLeuSerGluLysIleLysLeuGluCysGlnProGluLeu 308
| | | | |
Db 121 GAAG 180
| | | | |
QY 309 SerGluThrSerGlnThrLeuProProLysProPheSerCysGlyArgSerGlyLysGly 328
| | | | |
Db 181 TCGAGACATCCAGAGCTCTGCTCCAGAGCTCTTCAATGTGGCGAGTGAAGAGGA 240
| | | | |
QY 329 HisLysArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLeuAla 348
| | | | |
Db 241 CATTAAGAGAAATCCCATTTGGAAGTACAGAAAGAAAGACTCTGTTAAAAAGCTGGCT 300
| | | | |
QY 349 ProGluPheSerLysValLysThrLysThrProLysHisSerProLysGluGluPro 368
| | | | |
Db 301 CCTGAATTTTCAAAAGTCAAAACAAACAAACCTCAAGAGCACTCTTATTAAAGAGGAACC 360
| | | | |
QY 369 CysGlySerLeuSerGluThrValCysLysArgGluLeuArgSerGlnGluThrProGlu 388
| | | | |
Db 361 TGTGGTTCCTTACTGTAACCTGTTGTAAACGTAATGTAGAGACCAAGAAACCCAGAA 420
| | | | |
QY 389 LysProArgSerSerValAspThrProProArgLeuSerThrProGluLysGlyProSer 408
| | | | |
Db 421 AAGCCCCGCTCTTCAAGTGAACACCCACCAAGACTCTCCATCCCAAAAGGAGCCAGC 480
| | | | |
QY 409 ThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuProTyrLeuSerThr 428
| | | | |
Db 481 ACCATCCCAAGAGAGAAAGCTTCTCAAGTAGATGAAATTTGCCGTAACCTTCCACC 540
| | | | |
QY 429 ThrGluMetTyrLeuCysArgTTPHisGlnProProProSerProLeuProLeuArgGlu 448
| | | | |
Db 541 ACGAAATGATATTGTGTCTGTGGACACACCTCCCATCCATCAGCTTATTAAGAGGAA 600
| | | | |
QY 449 SerSerPro 451
| | | | |
Db 601 TCCTTCTCA 609
| | | | |
RESULT 14
| | | | |
LOCUS A1800794 549 bp mRNA linear EST 19-DEC-1999
| | | | |
DEFINITION W613C04.x1 Soares NSF P8 9W OT PA P S1 Homo sapiens cDNA clone
| | | | |
IMAGE:2364966.3' similar to contains MER22.b3 MER22 repetitive
| | | | |
element; mRNA sequence.
| | | | |
ACCESSION A1800794
| | | | |
VERSION A1800794.1 GI:5366266
| | | | |
KEYWORDS EST.
| | | | |
SOURCE Homo sapiens (human)
| | | | |
ORGANISM Homo sapiens
| | | | |
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
| | | | |
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
| | | | |
1 (bases 1 to 549)
| | | | |
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE	National Cancer Institute,Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaaps-remail.nih.gov This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert length: 1197 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 451.
FEATURES	location/Qualifiers
source	1..549 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:2364966" /lab_host="MDH10B" /clone_lib="Soares NSF P8 9M OT PA P S1" /note="Organ: pooled; Vector: pTf7D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NBHP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HF8-9M pool 1: 758980-760583 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 735080-74099 Subtraction by Bento Soares and M. Fatima Bernaldo."
ORIGIN	
Alignment Scores:	
Pred. NO.:	8.2e-158
Score:	178.00
Percent Similarity:	100.00%
Best local Similarity:	100.00%
Query Match:	28.9%
DB:	1 Gaps: 0
US-10-054-935-2 (1-614)	x AI800794 (1-549)
Gy	11 A1aa1aa1aProiaG1gYlVAsnProgiugInargileuAspiYrgIuaGAla1a1a 30
Dd	1 GCCGGCCCTCCTGCCGGGCAGACHTCTAGACGCAATTGAGACTAGAGCGGCTGCCG 60
Gy	31 LeuGIyglYProgiuaspGluProGiYAla1a1agiualah1apheLeuProArHiSaYg 50
Dd	61 CTGGGGCGGGCCGAGAGACGCTGGGGCGGGCCGAAGCCCACTTCCTCCCCGGGACCGT 120
Gy	51 LysLeuYsgIuPProgiYProProleuAlaSerSergIngiYglYserProAlaPProser 70
Dd	121 AAGCTCAAGAGCGCGGGCCCCCGGTGGCTCTCCCAAGCGGGGAGCCCCGGCCCTTCC 180
Gy	71 ProIlaagIcySgIyglYlVasGlYARgIyleuLeuProlaagiYAla1a1aProgiY 90
Dd	181 CCGGCGGGCTCGGGGCGAGGAGGGCGGGGGCTTTGTACTCCGAGCGGGGGGAGCCCCCGG 240
Gy	91 GlnGlnugiugIuseTerTrgiYglYserValProleuPProCySProProProAlaThrLyS 110
Dd	241 CAGCAGAGAAGAGACTGGGGGGGCTTGCGTGCCTTGGCCCTTGCCTCCGCCCGGCCACCAAG 300
Gy	111 GlnaIaagiYllegIyglYgluPProAla1a1aagiYalagiCYseSePProArpProlys 130
Dd	301 CAAGCCGGCATTTGGGGGGGAGCTTCGCGACGCGGAGCGGCTGAGCCCCCGGCCAAG 360
Gy	131 TyrGlnaVaAllauPProIleGIInThrclySerLeuVala1a1a1a1a1yGluPProthr 150
Dd	361 TATCAGGGGGGTCGCCCATTTCAAGCGGGCTCTCTCGGGGGGGCGGCGCAAAGAGCTTAGC 420

QY	151	Prothrlaplga1yapblysg1ygl1yala1aseProthralalathr1alaserAapPpOLA	170
Db	421	CCCTGGGCTGGGGCAACAGGGGTGGGGCCGCTCCCTCCCTCCACCGCTTCGGACCCGGCG	480
QY	171	Gl1yProBProleuPProleuPProgl1yProProleu1a1aProthralathr	188
Db	481	GGACCCCCACACTACTCTGCGCGGGCCGCCACCCCTCGCGCCACCGGCACC	534
RESULT 15			
LOCUS	CA444700/c		
DEFINITION	UI-H-DH1-aw-h-01-0-UI.s1 NCI CGAP DH1 Homo sapiens	linear	EST 08-NOV-2002
ACCESSION	CA444700		
VERSION	UI-H-DH1-aw-h-01-0-UI 3', mRNA sequence.		
KEYWORDS	CA444700.1	GI:24809120	
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
COMMENT	Tumor Gene Index		
	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: Dr. Jose Mercuende		
	CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa		
	CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa		
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa		
	Clone Distribution: Clone distribution information can be obtained		
	from Dr. M. Bento Soares, bento-soaresuiowa.edu		
	Seq primer: M13 FORWARD		
	POLYA=Yes.		
FEATURES			
source	location/Qualifiers		
	1..696		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="UI-H-DH1-aw-h-01-0-UI"		
	/tissue_type="Metastatic Chondrosarcoma"		
	/dev_stage="Adult"		
	/lab_host="DH10B (Life Technologies)"		
	/clone_lib="NCI CGAP DH1"		
	/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a		
	modified polylinker; Site_1: Ecor I; Site_2: Not I;		
	NCI CGAP DH1 is a normalized cDNA library containing the		
	following tissue(s): VS-8 Cell line from Metastatic		
	Chondrosarcoma in lung. The library was constructed		
	according to Bonaldo, Lennon and Soares, Genome Research,		
	6:791-806, 1996. First strand cDNA synthesis was primed		
	with an oligo-dT primer containing a Not I site. Double		
	stranded cDNA was ligated to an Ecor I adaptor, digested		
	with Not I, and cloned directionally into pT73-Pac		
	vector. The oligonucleotide used to prime the synthesis of		
	first-strand cDNA contains a library tag sequence that is		
	located between the Not I site and the (dT)18 tail. The		
	sequence tag for this library is AGATCATTCG.		
	TAG_TISSUE=lung		
	TAG_LIB=UI-H-DH1		
	TAG_SEQ=AGATCATTCG"		
ALIGNMENT SCORES:			
Pred. No.:	7.2e-155	Length:	696
Score:	175.00	Matches:	175
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	28.50%	Indels:	0
DB:	6	Gaps:	0

US-10-054-935-2 (1-614) x CA444700 (1-696)

QY 284 GlnGlyTyrGluThrGluArgGluGluThrGluLeuSerGluYsIleYsLeuGlu 303
 Db 658 CAGGGCTATGAACTGAAGAGAGAGAAACAGAGCTATCTGAAATTAATTAACCTGAG 599
 QY 304 CysGlnProGluLeuSerGluThrSerGlnThrLeuProProlyProPheSerCysGly 323
 Db 598 TGCAGCGGAGCTTCCGAGACATCCAGACTGCTGCCCTCCCAAGCCCTTCTCATGTGGG 539
 QY 324 ArgSerGlyYsGlyYsIleYsArgYsSerProPheGlySerThrGluArgYsLeuThrPro 343
 Db 538 CGAGGTGAAAAGGACATTAAGAGAAATCCCATTTGTAAGTACAGAAAGAAACATCTCT 479
 QY 344 ValIysIleLeuAlaProGluPheSerYsValIysThrYsThrProYsHisSerPro 363
 Db 478 GTTAAAGAGTGGCTCTGTAATTTTCAAAAGTCAAAACAAACTCTTAAGCACTTCCT 419
 QY 364 IleYsGluGluProCysGlySerLeuSerGluThrValCysYsArgGluLeuArgSer 383
 Db 418 ATTAAGAGAACCTCTGCTTCTTATCTGAACCTGTTGTAAACGTGAATTGAGAGGC 359
 QY 384 GlnGluThrProGluYsProArgSerSerValAspThrProProAlaGluSerThrPro 403
 Db 358 CAAGAAACCCCAAGAAAGCCCGGTCTTCAGTGAACCCCAACAGACTCTCCACTCC 299
 QY 404 GlnYsGlyProSerThrIleProYsGluYsAlaPheSerSerGluIleGluAspLeu 423
 Db 298 CAAGAGGAGACCCAGACCCATCCCAAGAGAAAGCTTCTCAAGTGAATGAAAGATTG 239
 QY 424 ProTyrLeuSerThrThrGluMetYrLeuCysArgTrpHisGlnProProSerPro 443
 Db 238 CCGTACCTTTCCACACAGAAATGATTTGTGTGGACCAAGCCCTCCCATCAACG 179
 QY 444 LeuProLeuArgGluSerSerProYsGlyGluGluThrValAla 458
 Db 178 TTACCATTAACGGAAATCTCTTCCAAAGAGAGAGAGACTGTAGCA 134

RESULT 16
 CK300608/c
 LOCUS 721 bp mRNA linear EST 15-DEC-2003
 DEFINITION UI-E-EJ1-a-j2-b-15-0-UI-s1 UI-E-EJ1 Homo sapiens cDNA clone
 ACCESSION CK300608
 VERSION CK300608
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 721)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL Genome Res. 6 (9) , 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Greg Hageman
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
 http://genome.uiowa.edu/distribution/eye.html
 Seq primer: M13 FORWARD
 POLYA=yes
 location/Qualifiers

SOURCE

1. 721
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-EJ1-a-j2-b-15-0-UI"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina foveal and Macular, RPE and
 Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_1lb="UI-E-EJ1"
 /note="Organ: eye; Vector: pRTT3-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-EJ1 is a subtracted cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pRTT3-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (drr)18 tail. The
 sequence tags for this library are: fetal eyes,
 AGATCAAGA; lens, CGATTAGCA; eye anterior segment,
 AATGCCGAT; optic nerve, CCATTAACTG; retina, CCGCG;
 foveal and macular, GTCC; RPE and Choroid, ACCTA. This
 library was created for the program, Gene Discovery in the
 Visual System, supported by National Eye Institute (NEI).
 TAG LIB=UI-E-EJ1
 TAG_SEQ=AGATCAAGA"

ORIGIN

Alignment Scores:

Pred. No.: 7,45e-155 Length: 721
 Score: 175.00 Matches: 175
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 28,50% Indels: 0
 DB: 7 Gaps: 0

US-10-054-935-2 (1-614) x CK300608 (1-721)

QY 284 GlnGlyTyrGluThrGluArgGluGluThrGluLeuSerGluYsIleYsLeuGlu 303
 Db 658 CAGGGCTATGAACTGAAGAGAGAGAAACAGAGCTATCTGAAATTAATTAACCTGAG 599
 QY 304 CysGlnProGluLeuSerGluThrSerGlnThrLeuProProlyProPheSerCysGly 323
 Db 598 TGCAGCGGAGCTTCCGAGACATCCAGACTGCTGCCCTCCCAAGCCCTTCTCATGTGGG 539
 QY 324 ArgSerGlyYsGlyYsIleYsArgYsSerProPheGlySerThrGluArgYsLeuThrPro 343
 Db 538 CGAGGTGAAAAGGACATTAAGAGAAATCCCATTTGTAAGTACAGAAAGAAACATCTCT 479
 QY 344 ValIysIleLeuAlaProGluPheSerYsValIysThrYsThrProYsHisSerPro 363
 Db 478 GTTAAAGAGTGGCTCTGTAATTTTCAAAAGTCAAAACAAACTCTTAAGCACTTCCT 419
 QY 364 IleYsGluGluProCysGlySerLeuSerGluThrValCysYsArgGluLeuArgSer 383
 Db 418 ATTAAGAGAACCTCTGCTTCTTATCTGAACCTGTTGTAAACGTGAATTGAGAGGC 359
 QY 384 GlnGluThrProGluYsProArgSerSerValAspThrProProAlaGluSerThrPro 403
 Db 358 CAAGAAACCCCAAGAAAGCCCGGTCTTCAGTGAACCCCAACAGACTCTCCACTCC 299
 QY 404 GlnYsGlyProSerThrIleProYsGluYsAlaPheSerSerGluIleGluAspLeu 423
 Db 298 CAAGAGGAGACCCAGACCCATCCCAAGAGAAAGCTTCTCAAGTGAATGAAAGATTG 239
 QY 424 ProTyrLeuSerThrThrGluMetYrLeuCysArgTrpHisGlnProProSerPro 443

Db 361 CAGATTCAGGGGCTCTGCCCATTTCAGACGGGCTCTCTGCGCGGCGCCAAAGGCC 420
 QY 149 oThrProTPrAlaGlyAspLys-GlyGlyAlaAlaSerProAlaAlaThrAlaSerAsp 169
 Db 421 TACGCCCTGGGCTGGGAGACAGAGGGGTGGGGCGGCTCTCCCGCTGCACCGCTCGGACC 480
 QY 169 tAlaGlyProPProPProLeuProLeuProGlyProPProPProLeuAlaProThAlaThra 189
 Db 481 CGCGGGAGCCCCACCACTACCTCTGCGCGCGCCACCCCTCTGCGCCACCGCACCG 540
 QY 189 lAglyThreulAlaAlaSerGluGlyArgTrpLysSerMet 202
 Db 541 CCGGAGCCTGGCGCGCAGGAGGAGATGAGAGTATG 581
 RESULT 20
 BQ711779 985 bp mRNA linear EST 16-JUL-2002
 LOCUS AGENCOURT_8487897 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6300573
 DEFINITION 5', mRNA sequence.
 ACCESSION BQ711779
 VERSION BQ711779.1 GI:21850678
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 985)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strauberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Dr. Mark Watson
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LNCM2515 row: a column: 22
 High quality sequence stop: 592.
 Location/Qualifiers
 1..985
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6300573"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 Laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 3.86e-147 Length: 985
 Score: 167.00 Matches: 167
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 27.20% Indels: 0
 Db: 5 Gaps: 0

US-10-054-935-2 (1-614) x BQ711779 (1-985)

QY 165 ThrlAlaSerAspProAlaGlyProPProPProLeuProLeuProGlyProPProPProLeuAla 184
 Db 11 ACCGCTCGAGCCCGGCGGAGCCCACTACCTCTGCGCGCGCCACCCCTCGCG 70

QY 185 ProThAlaThrAlaGlyThreulAlaAlaSerGluGlyArgTrpLysSerMetArgLys 204
 Db 71 CCACACGCCACCGCCCGGACCTTGCGGCGCACGAGAGGAGATGAGAGTATGAGAGG 130
 QY 205 SerProLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 224
 Db 131 AGCCCTCTGGGGGTGGTGGCGGCTCGGAGCTCCAGTCAAGCCCGCTCTCCAAACAG 190
 QY 225 lLeuLeuLeuGluLeuAspLeuileglinginglinglinglinglinglingling 244
 Db 191 ATCTTCTCTGCAATTTGACCTCATCGAAGACAGACAGACAGACAGAGCTGCAGCGCAAGAA 250
 QY 245 lYsglIuileglinglinglinglinglinglinglinglinglinglinglingling 264
 Db 251 AAGGATGAGAGAGCTGAAGTCAAGAGAGACACAGCTCTGCTGCTGGAATGAACCTATG 310
 QY 265 GluArgArgMetGluLeuVallylYslyAspAspGluGlyGlyGlyGlyGlyGlyGly 284
 Db 311 GAAAGCGGATGCGAGCTGTAAAGAGATACGAGAAAGAAAGGACAGAGCTGTTTCAG 370
 QY 285 GlyTyrgluThrgluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 304
 Db 371 GGCTATGAACCTGAAG 430
 QY 305 GluProGluLeuSerGluThrSerGluThrLeuProProPProPProPProPProPPro 324
 Db 431 CAGCCGAGCTTCCAGACATCCAGACTCTGCTCCCAAGCCCTTTCATGATGGGCGG 490
 QY 325 SerGlyLysGlyYHisLysArg 331
 Db 491 AGTGAAAGGAGCATTAAGG 511

RESULT 21
 LOCUS Bx478274 560 bp mRNA linear EST 04-SEP-2003
 DEFINITION DKFZp686L05203 r1 686 (synonym: h1cc3) Homo sapiens cDNA clone
 Bx478274
 ACCESSION Bx478274
 VERSION Bx478274.1 GI:31912944
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 560)
 AUTHORS Ansoerge, W., Krieger, S., Regiert, T., Rittmuller, C., Schwager, B.,
 Mewes, H.W., Weill, B., Amlid, C., Oesinger, A., Fobo, G., Han, M. and
 Wiemann, S.
 EST (Ansoerge, W., Krieger, S., Regiert, T., Rittmuller, C., et al.)
 Unpublished (2003)
 COMMENT Contact: MIPS

TITLE Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
 JOURNAL This is the 5' sequence of the clone insert
 COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;
 sequenced by EMBL (European Molecular Biology Laboratories,
 Heidelberg/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 No st sequence available.
 This clone (DKFZp686L05203) is available at the RZPD in Berlin.
 Please contact the RZPD: Reesourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY, Email: clone@rzpd.de.

FEATURES

source
 1..560
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp686L05203"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="686 (synonym: h1cc3)"

ORIGIN

/note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

Alignment Scores:

Pred. No.: 1.75e-145 Length: 560
Score: 165.00 Matches: 165
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.87% Indels: 0
DB: Gaps: 0

US-10-054-935-2 (1-614) x BX478274 (1-560)

Qy 344 VallylsylsleuAlaProgluPheserlyValylserThyThrProlyGHisSerPro 363
Db 51 GTTAAAGAGTGGCTCTGAAATTTTCAAAAGTCAAAACAAAGTCTCTAGACCTCTCT 110
Qy 364 IleySGluGluProCySGlySerleuSerGluThrValCyslySArgGluLeuArgSer 383
Db 111 ATTAAGAGGAAACCTGCTGCTTCTTATCTGAGACTGTTGTAAACGTAATTGAGAGC 170
Qy 384 GluGluThrProGluThrProArgSerSerValAspThrProProArgLeuSerThrPro 403
Db 171 CAAGAAACCCAGAAAGCCCGGCTTTCAGTGGACACCCACCAAGACTCTCCACTCCC 230
Qy 404 GluHlySGlyProSerThrHisProlyGluValylAspSerSerGluIleGluAspLeu 423
Db 231 CAAGAGGAGCCAGACCCAGTCTTCAAGAGAGAAAGCTTCAAGTGAATGAAGATTGG 290
Qy 424 ProTyIleuSerThrThrGluMetTyIleuCySArgTrpHisGlnProProSerPro 443
Db 291 CCGACTTTCACCAACCAAGAAATGATTGTGTGCTGGACCAAGCTCCCAATCACCG 350
Qy 444 LeuProleuArgGluSerSerProlyGlyGluGluThrValAlaArgCysLeuMetPro 463
Db 351 TTACCATTAAGGAGGAAATCTCTCCAAAGAGAGAGAGACTGTGATGATGCCA 410
Qy 464 SerSerValAlaGlyGluThrSerValLeuAlaValProSerTPArgAspHisSerVal 483
Db 411 TCAAGTGTGAGAGAAACTTCAGTCTTGGCTTTCCTTTCGAGGAGACACTCAGTA 470
Qy 484 GluProleuArgAspProAsnProSerAspLeuGluLeuAspAspSerValPhe 503
Db 471 GAGCTTAAGGAGCCAAATCTTCAAGACTTTTGAAGAACTGGATGACAGTGTGT 530
Qy 504 SerlySArgHisAla 508
Db 531 TCGAAGCGCATGCA 545

RESULT 22
BQ127983 508 bp mRNA linear EST 15-JUL-2003
LOCUS 1179908.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:5777847
DEFINITION BQ127983
5', mRNA sequence.
ACCESSION BQ127983
VERSION BQ127983.1 GI:20201894
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 508)
AUTHORS Melton,D., Brown,J., Kenly,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Scaerce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistean,A.,
Schmitt,A., Theising,B., Ritter,B., Ronko,I., Bennett,J.,
Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarashvili,R.,
Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium

FEATURES

source
1..508
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5777847"
/issue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="Human insulinoma"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
XhoI; Site_2: EcoRI; Constructed with lambda ZapII system
(Stratagene) by Dr. J. Ferrer. In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."

ORIGIN

Alignment Scores:

Pred. No.: 1.25e-143 Length: 508
Score: 163.00 Matches: 163
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.55% Indels: 0
DB: Gaps: 0

US-10-054-935-2 (1-614) x BQ127983 (1-508)

Qy 189 AlAGlyThrleuAlaAspSerGluGlyArgTrpIleSerMetArglySerProleuGly 208
Db 18 GCCGGAGCCCTGGCGGCAGCGAGGAGAGATGAGATGAGAGAGAGAGAGAGAGAG 77
Qy 209 GlyGlyGlySerGlyAlaSerSerGlnAlaAlaCysleuGlyGlnIleleuLeu 228
Db 78 GGTGGTGGCGGCTCGGAGGCTCTCAGTCAAGCGCCGCTCCAAACAGATCTTCTGCTG 137
Qy 229 GluLeuAspLeuIleGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 248
Db 138 CAATTGAGCTCATCGAAG 197
Qy 249 GluLeuAspSerGluArgAspThrleuLeuAlaArgIleGluArgMetGluArgMet 268
Db 198 GAGCTGAAGTCAGATG 257
Qy 269 GluLeuValIleValAspAspGluValGlyArgHisIleValAspLeuPheGlnGlyThr 288
Db 258 CAGCTGTAAGAGAGATTAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 317
Qy 289 GluGluArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 308
Db 318 GAAG 377
Qy 309 SerGluThrSerGlnThrIleuProProlySerProPheSerCysGlyArgSerGlyGly 328
Db 378 TCCGAGACATCCAGACTCTGCTCCCAAGCCCTTTCAGTGGCGGAGAGAGAGAGAGAG 437
Qy 329 HisIleArglySerProPheGlySerThrGluArgThrProValIleValLeuAla 348

Db 438 CATTAAGGAAATCCCATTTGGAAGTACAGAAAGAAAGACTCTGTATAAAAGCTGGCT 497
 QY 349 ProgluPhe 351
 Db 498 CCGAATTT 506

RESULT 23
 CA307817/c 640 bp mRNA linear EST 05-AUG-2004
 LOCUS
 DEFINITION
 UI-H-FTI-bhx-d-08-0-UI.s1 NCI CGAP FTI Homo sapiens CDNA clone
 CA307817
 CA307817.1 GI:24470871
 EST.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 640)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bcrfemail.nih.gov
 Tissue Procurement: Dr. Gary W. Hunninghake, U of I
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@uiowa.edu
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES
 source
 1. .640
 /location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-FTI-bhx-d-08-0-UI"
 /tissue_type="Alveolar Macrophage"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_idb="NCI_CGAP_FTI1"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site: 1: Ecor I; Site 2: Not I;
 NCI CGAP FTI 1 is a normalized CDNA library constructed from
 a pool of 81 RNA samples from Alveolar Macrophages
 challenged with different treatments. The RNA samples
 were a mixture of these conditions (times refer to
 incubations following isolation by bronchoalveolar lavage)
 (some normal donor macrophages were cultured in some of
 the conditions, other donor macrophages in different
 conditions). The mRNA samples were pooled for library
 construction. Control 0 hours; control 3 hours; control 24
 hours; LPS 100 ng/mL, 3 hours; LPS 100 ng/mL, 24 hours;
 PMA 10 ng/mL, 3 hours; PMA 10 ng/mL, 24 hours; Klebsiella
 moi 10, 3 hours; Klebsiella moi 10, 24 hours; Staph aureus
 moi 10, 3 hours; Staph aureus moi 10, 24 hours; Adenoviral
 vector (Ad5 CMV egfp), moi 500, 3 hours; Adenoviral vector
 (Ad5 CMV egfp), moi 500, 24 hours; wt adenovirus moi 500,
 3 hours; wt adenovirus moi 500, 24 hours; Ad vector + LPS
 3 hours; Ad vector + LPS 24 hours; The library was
 normalized according to Bonaldo, Lennon and Soares, Genome
 Research, 6:791-806, 1996. First strand cDNA synthesis was
 primed with an oligo-dt primer containing a Not I site.
 Double stranded cDNA was ligated to an Ecor I adaptor,
 digested with Not I, and cloned directionally into
 pT73-Pac vector. The oligonucleotide used to prime the
 synthesis of first-strand cDNA contains the Not I site and the
 sequence that is located between the Not I site and the
 (dt)18 tail. The sequence tag for this library is

ORIGIN
 Alignment Scores:
 Pred. No.: 1.39e-142 Length: 640
 Score: 162.00 Matches: 162
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 26.38% Indels: 0
 DB: 6 Gaps: 0

US-10-054-935-2 (1-614) x CA307817 (1-640)
 QY 453 LysGluGluThrValAlaArgCysLeuMetProSerSerValAlaGlyGluThrSerVal 472
 Db 638 AAGAGAGAGACTGTACCAAGGTGTCTGATGCCATTCAGTGTTCAGAGAACTTCAGTC 579
 QY 473 LeuAlaValProSerTrpArgAspHisSerValGluProLeuArgAspProAsnProSer 492
 Db 578 TTGGCTGTTCCTCTTGGAGGAGCACTCAGTAGACCTCTTAAGGACCAATCTTCA 519
 QY 493 AspleuLeuGluAsnLeuAspAspSerValPheSerLysArgHisAlaLysLeuGluLeu 512
 Db 518 GACCTTTTGGAGAACTGATGACAGTGTGTTTTCACACCGCATCAAACTGAGAGCTG 459
 QY 513 AspGluLysArgArgLysArgTrpAspIleGlnArgIleArgGluGluGln 532
 Db 458 GATGAGAAAGAGAGAAAGAGATGAGATTCACAGAGATCAGGAAACAAATTTTACG 359
 QY 533 ArgLeuGluLeuArgMetTrpLysLysLysGlyIleGlnGlnSerGluProGluValThr 552
 Db 398 CGACTCGAGCTCAGATGTATATAAAGAAAGAAATTCAGAAATCTGAGCTGAGTTACC 339
 QY 553 SerPhePheProGluProAspAspValGluSerLeuMetLethProPheLeuProVal 572
 Db 338 TCAATTTTCCCTGAGCAGATGATGTTGAAGTTTATGATGATCCCTTCTGCTGTT 279
 QY 573 ValAlaPheGlyArgProLeuProLysLeuThrProGluAsnPheGluLeuProTyrPleu 592
 Db 278 GTTGCATTTTGGACGACCATTTACCAAAATTTACTCCAGAAATTTTACTCTGCTTGG 219
 QY 593 AspGluArgSerArgCysArgLeuGluIleGlnLysLysGlnThrProHisArgThrCys 612
 Db 218 GATGACGATGACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 159
 QY 613 ArgLys 614
 Db 158 AGGAAA 153

RESULT 24
 BX952313 644 bp mRNA linear EST 01-MAR-2004
 LOCUS
 DEFINITION
 DXFZp781M12183 r1 781 (synonym: h1cc4) Homo sapiens CDNA clone
 BX952313
 BX952313.1 GI:43428943
 EST.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 644)
 Mamburt, R., Heubner, D., Mewes, H. W., Weill, B., Amid, C., Osanger, A.,
 Fobbo, G., Han, W., and Wiemann, S.
 EST (Mamburt, R., Heubner, D., Mewes, H. W., Weill, B., Amid, C., et al.)
 Unpublished (2003)
 CONTACT: MIPS
 MIPS
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

pcmv-SPORT6 plasmid vector. The average insert size is about 3.6kb."

ORIGIN

Alignment Scores:

Pred. No.:	7.36e-136	Length:	789
Score:	155.00	Matches:	155
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	25.24%	Indels:	0
DB:	6	Gaps:	0

US-10-054-935-2 (1-614) x CD652041 (1-789)

```
QY 460 CysLeuMetProSerSerValAlaGlyGlnThrSerValLeuAlaValProSerTyrArg 479
    |||
Db 107 TGTCTATGTCATCAAGTGTTCAGAGAACTTCAGTCTTGGTGTCTTCTTGAGG 166
QY 480 AspHisSerValGluProLeuArgAspProAsnProSerAspLeuGluAlaGlnLeuAsp 499
    |||
Db 167 GACCACTCACTAAGCCTCTAAGGACCCAAATCCTTCAGACCTTTGGAGAACCTGGAT 226
QY 500 AspSerValPheSerLysArgHisAlaAlaLysLeuGluLeuAspGluLysArgArgLysArg 519
    |||
Db 227 GACAGTGTGTTTCGAAGCCGCAATGCAAACTGGAGCTGGATGAGAGAGAGAAAGAA 286
QY 520 TrpAspTleGlnArgTleArgGluGlnArgTleLeuGlnArgLeuGlnLeuMetTyr 539
    |||
Db 287 TGGGATATTCAGAGGATCAGAGGAAACAAGATTTTACAGCACTGCGCTCAGATATAT 346
QY 540 LysLysLysGlyTleGlnGlnSerGluProGluValThrSerPhePheProGluProAsp 559
    |||
Db 347 AAAAAGAAAGAAATTCAGAAATCTGAGCTGAGGTTTCTCATTTTCTTCCAGCCAGAT 406
QY 560 AspValGluSerLeuMetTleThrProPheLeuProValAlaPheGlyArgProLeu 579
    |||
Db 407 GATGTGAAGATTGATGATTACCCCTTCTGCTGTGTAGCATTTGGACACCATTA 466
QY 580 ProLysLeuThrProGlnAsnPheGluLeuProTTrpLeuAspGluArgSerArgCysArg 599
    |||
Db 467 CCAAAATTAATCTCACAGAAATTTGAGCTACCCCTGTTGATGATGACCTAGCCGATGAGA 526
QY 600 LeuGlnTleGlnLysLysGlnThrProHisArgThrCysArgLys 614
    |||
Db 527 TTGGAGATCCAGAGAAACAACCTCACCGGACGTTAGAGAA 571

RESULT 26
AI954139 463 bp mRNA linear EST 06-SEP-1999
LOCUS wx80B09.x1 NCI_CGAP_Ov38 Homo sapiens cDNA clone IMAGE:254945 3',
DEFINITION mRNA sequence.
ACCESSION AI954139
VERSION AI954139.1 GI:5746449
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 463)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeff Struwing, M.D., Michael R. Emmert-Buck,
M.D., Ph.D. cDNA library Preparation: Life Technologies, Inc. cDNA
library Arrayed by: Christa Frange, The I.M.A.G.E. Consortium DNA
Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bdrr/image/image.html
Seq primer: -40UP from Gidco
```

High quality sequence stop: 447.
Location/Qualifiers
1.463

FEATURES
source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:254945"
/issue_type="normal epithelium"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Ov38"
/note="Organ: ovary; Vector: pcmv-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: 01lgo dr.
Library constructed by Life Technologies."

ORIGIN

Alignment Scores:

Pred. No.:	3.89e-135	Length:	463
Score:	154.00	Matches:	154
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	25.08%	Indels:	0
DB:	1	Gaps:	0

US-10-054-935-2 (1-614) x AI954139 (1-463)

```
QY 10 AlaAlaAlaAlaProAlaGlyGlyAsnProGlnArgLeuAspTyrGluArgAla 29
    |||
Db 2 GCGCCGCGGCCCCCTCCGCGGCAATCCTGAGCAGACTGACTACGAGCGGGCTGGC 61
QY 30 AlaLeuGlyGlyProGluAspGluProGlyAlaAlaGlyAlaHisPheLeuProArgHis 49
    |||
Db 62 GCGCTGGCGCGGCCAGAGACAGACCTGGGGCGCGCGGAAGCCACTTCTCCCGGCGAC 121
QY 50 ArgLysLeuLysGluProGlyProProLeuAlaSerSerGlnGlyGlySerProAlaPro 69
    |||
Db 122 CGTAAGCTCAAGAGAGCGGGGCCCGCGCTGCTCTCCAGAGGCGGAGCCCGGCGCT 181
QY 70 SerProAlaGlyCysGlyGlyLysGlyArgGlyLysLeuLeuLeuProAlaGlyAlaAlaPro 89
    |||
Db 182 TCCCGCGCGCGCGCGGCGGCAAGGGCGGGGCTTTTACTCCCGCGGGGCGGCCCC 241
QY 90 GtLgngLgngLgusertPgiYgYservalProLeuProCysPProProAlaThr 109
    |||
Db 242 GGGCAGCAGAGAGAGAGCTGGGGCGGTTCGTGCTTCCCTGTCGCCCCGCGCAC 301
QY 110 LysGlnAlaGlyTleGlyGluProAlaAlaAlaGlyAlaGlyCysSerProArgPro 129
    |||
Db 302 AACCAAGCCGGCAATGGGGGAGCCTGCGCGAGCGGAGCCGCTGCAAGCCCGGCGCC 361
QY 130 LysTyrGlnAlaValLeuProTleGlnThrGlySerLeuValAlaAlaLysGluPro 149
    |||
Db 362 AAGTATCAGCGGTGTGCTGCCATTGACAGCGGCTCTCTGTGGCGGCGCAAGAGCT 421
QY 150 ThrProTrpAlaGlyAspLysGlyGlyAlaAlaSerProAla 163
    |||
Db 422 ACGCCCTGGGCTGGGACAAAGGTGGGGCGGCTTCCCGGCT 463

RESULT 27
AI582361 460 bp mRNA linear EST 14-DEC-1999
LOCUS tq67P10.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2213851 3',
DEFINITION sequence.
ACCESSION AI582361
VERSION AI582361.1 GI:4568258
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 460)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```



```
Pred. No.: 3 88e-133 Length: 591
Score: 152.00 Matches: 152
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.76% Indels: 0
DB: 5 Gaps: 0
US-10-054-935-2 (1-614) x BM969790 (1-591)

QY 307 GluLeuSerGluThrSerGlnThrLeuProProlySerProheSerCysGlyArgSerGly 326
Db 589 GACCTTCCGAGACATCCAGACTCTGCTCCCAAGCCCTTCATGAGGCGAGTGA 530
QY 327 LysGlyHisLysArgLysSerProheGlySerThrGluArgLysThrProValLys 346
Db 529 AAGGACATAAAGAAATCCCATTTGAAAGTACAAAGAAAGACTCTGTTAAAG 470
QY 347 LeuAlProGluPheSerLysValLysThrLysThrProLysHisSerProLysGlu 366
Db 469 CTGGCTCTGATTTTCAAAAGTCAAAACAAACCTCTAAGCACTCTCTATTAAAGAG 410
QY 367 GluProCysGlySerLeuSerGluThrValCysLysArgLysLeuArgSerGlnGluThr 386
Db 409 GAACCTGTGGTTCCTTATCTGAACCTGTTGTAAAGCTGAATTGAGAGCCAAAGAAC 350
QY 387 ProGluLysProArgSerSerValAspThrProProArgLeuSerThrProGlnLysGly 406
Db 349 CCAAGAAAGCCCGGCTTCAGTGAGACCCCAAGACTCTCCATCCCAAAAGGA 290
QY 407 ProSerThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuProLysLeu 426
Db 289 CCAAGCAACCATCCCAAGAGAAAGCTTTCTCAAGTAAAGATTTGCCCTACTT 230
QY 427 SerThrThrGluMetLysLeuCySarGTPHisGlnProProSerProLeu 446
Db 229 TCACACACAGAAATGATTTGTGTGCTGGACACAGCCTCCCATCAGCTTACATTA 170
QY 447 ArgGluSerSerProLysLysGluGluThrValAla 458
Db 169 CGGGAATCCTCTCCAAAGAGAGAGACTGTAC 134

RESULT 29
BI457952 623 bp mRNA linear EST 21-AUG-2001
LOCUS DEFINITION 603198945F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5278546 5',
mRNA sequence.
BI457952
BI457952.1 GI:15248608
EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 623)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshitaki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHAM1703 row: i column: 11
High quality sequence stop: 623.
Location/Qualifiers
1. 623
/organism="Homo sapiens"
/mol_type="mRNA"
```

```
/db_xref="taxon:9606"
/clone="IMAGE:5278546"
/issue_type="hypothalamus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_96"
/notes="Organ: Brain; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI; XhoI
(gtcgag); Oligo-dT primed using primer
5'-TTTTTTTTTTTAA-3', size-selected for average
insert size 2.3 kb and normalized to R07 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: this is
a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 4 08e-133 Length: 623
Score: 152.00 Matches: 195
Percent Similarity: 98.98% Conservative: 0
Best Local Similarity: 98.98% Mismatches: 1
Query Match: 24.76% Indels: 2
DB: 4 Gaps: 0
US-10-054-935-2 (1-614) x BI457952 (1-623)

QY 263 ArgMetGluArgArgMetGlnLeuValLysLysAspAsnGluLysGluArgHisLysLeu 282
Db 3 CGAGTGAAGAAAGCGAGTGCAGTGGTAAAGAGATACAGAAAGAAAGGCAAGCTG 62
QY 283 PheGlnGlyTyrGluThrGluGluArgGluGluThrGluLeuSerGluLysLysLeu 302
Db 63 TTTCAGGGCTATGAACTAAAGAGAGAGAGAAACAGACTATCTGAATAATTTAACTG 122
QY 303 GluCysGlnPro-GluLeuSerGluThrSerGlnThrLeuProProLysProheSerCys 322
Db 123 GAATGCCAGCG-GGACTTTCCAGACATCCAGACTTGTCTCCAGCCCTTCTCATG 181
QY 322 sglYArgSerGlyLysGlyHisLysArgLysSerProheGlySerThrGluArgLysTh 342
Db 182 TGGCGGAGTGAAGAAAGGACATAAAGAAATCCCATTTGAAAGTACAGAAAGAAAGAC 241
QY 342 rProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysHis 362
Db 242 TCCTATTAAAGAAAGCTGCTCTGAATTTCAAAAGTCAAAACCAAACTCTTAAGCACTC 301
QY 362 rProLysGluGluLysProCysGlySerLeuSerGluThrValCysLysArgGluLeu 382
Db 302 TCCTATTAAAGAAAGAAAGCTGTGGTTCCTTATCTGAACCTGTTGTAACGTGAATTGAG 361
QY 382 gSerGlnGluThrProGluLysProArgSerSerValAspThrProProArgLeuSerTh 402
Db 362 GAGCCAAAGAAAGCCCAAGAAAGCCCGGCTTTCAGTGAACACCCCAAGACTCTCCAC 421
QY 402 rProGlnLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIleGlu 422
Db 422 TCCCAAAAGAGGAGCCAGACCCATCCCAAGAGGAAAGCTTCTCAAGTACATGAAGA 481
QY 422 pLeuProLysLeuSerThrThrGluMetLysLeuCySarGTPHisGlnProProhe 442
Db 482 TTGGCGTACCTTTCCACACAGAAATGATTTGTGTGCTGGACACAGCTCCCATC 541
QY 442 rProLeuProLeuArgLysSerSerProLysLysGluGluThrValAla 458
Db 542 ACCGTTACATTCAGGAAATCCTCTCCAAAGAGAGAGACTGTAC 590

RESULT 30
AI831057 451 bp mRNA linear EST 21-DEC-1999
LOCUS DEFINITION w162d01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407393 3',
mRNA sequence.
AI831057
```

```

VERSION      AI831057.1  GI:5451728
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE        Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgaaps-remail.nih.gov
              Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
              Emmer-Buck, M.D., Ph.D.
              CDNA Library Preparation: M. Bento Soares, Ph.D.
              CDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E.B. Consortium/LINL ac:
              www-bio.llnl.gov/bdip/image/image.html
              Insert Length: 2097 Std Error: 0.00
              Seq primer: -40UP from Gidpc
              High quality sequence atopc: 440.
FEATURES
  Source
    1..451
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:2407393"
    /tissue_type="squamous cell carcinoma, poorly
    differentiated (4 pooled tumors, including primary
    metastatic)"
    /dev_stage="adult"
    /lab_host="MDH10B (phage-resistant)"
    /clone_lib="NCI-CGAP_Lin9"
    /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
    modified polylinker; 1st strand cDNA was prepared from
    pooled lung tumor tissue, and was then primed with a Not I
    - oligo(dT) primer. Double-stranded cDNA was ligated to
    Eco RI adaptors (Pharmacia), digested with Not I and
    cloned into the Not I and Eco RI sites of the modified
    pT73 vector. Library went through one round of
    normalization. Library constructed by Bento Soares and M.
    Fatima Bonaldo."
ORIGIN
Alignment Scores:
Pred. No.:          2,35e-131           Length:         451
Score:             150.00               Matches:         150
Percent Similarity: 100.00%            Conservative:    0
Best Local Similarity: 100.00%          Mismatches:     0
Query Match:       24.43%              Indels:         0
                               Gaps:         0
US-10-054-935-2 (1-614) x AI831057 (1-451)
QY      10 A1A1AA1AA1AProIAAGLYVAsnProGIUGInAtgtLeuAsPTyTG1AArG1AA1A 29
Db      1 GCGGCCGGCGGCCCTGTCCGGCGGCACATCTCAGAAGCACTGGACTACGAGCCGGGCTGCG 60
QY      30 A1A1euglyg1yPProGL1uAsPG1uPProGL1A1A1A1Agi1ua1A1A1A1A1A1A1A1A 49
Db      61 GGCGTGGCGGGGCCCGAGAGACGAGCTGGGGGGGGGGCGGAGGCCCATCTTCCTCCCGGGCAC 120
QY      50 Arg1ys1eu1ygs1uProGL1yPProFroLeu1AasErsGing1ygl1SerProAlabRo 69
Db      121 CGTAAGACTCAAGAGAGCCGGGGCCCCCGCTGGCTCTCTCCAGAGGGGGGAGCCCCGGGCTT 180
QY      70 SerProIAagi1yCyBg1yGL1yUygs1yArGL1yLeu1eu1euProIAagi1yA1AA1APro 89
Db      181 TTCCTGGCGGCGCTGGCGGCGGCAAGGGCGGGGCTTTGATTCTCCGGCGGGGGCGGCCCCCCC 240

```

Qy	90	gylglnlgnlunlueertfpglygylgylsevalproleuprocysppropioalathr	109
Db	241	gggcacgacgaagagacgctggggcggttcgggtgcttccgcttccgccccggccacc	300
Qy	110	lysglnalaglylleglygylguiproalaaalalaglylcysserproatrgpro	129
Db	301	aagcaagccggcatttgggggggagcctgcgcgacagccgagccggactcagccccggacc	360
Qy	130	lyetrglnalalvalleuprollegrnthrglyserleuvalalalalalygluipro	149
Db	361	aaatcatcagcggctgctgccttcacacgagcgctctctctggcgcgccgacaaagacct	420
Qy	150	thpctrrpalaglyasplysglygylala	159
Db	421	acgcccttgagctggcgacaaaggctggagcg	450
RESULT 31			
LOCUS	AM854130	501 bp	mRNA linear EST 19-MAY-2000
DEFINITION	RC3-CT0254-060400-029-a11 CT0254 Homo sapiens cDNA, mRNA sequence.		
ACCESSION	AM854130		
VERSION	AM854130.1	GI:7949823	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 501) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsuoka,A., Bata,G.S., Simpson,D.H., Brundel,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
MEDLINE	20202663		
PUBMED	10737800		
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=kt2-RC3-CT0254-060400-029-a11&t3=2000-04-06&t4=1) Seq primer: puc 18 forward High quality sequence stop: 501. Location/Qualifiers 1..501 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_id="CT0254" /note="Organ: colon; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORSRES PCR (U.S. Letters Patent application No. 196,776 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
FEATURES			
SOURCE			
ORIGIN			
Alignment Scores:			
Pred. No.:	2.6e-131	Length:	501
Score:	150.00	Matches:	150
Percent Similarity:	100.00%	Conservative:	0

On 286 Twelfth Avenue

Percent Similarity:	149.00	Matches:	149
Score:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	24.27%	Indels:	0
DB:	7	Gaps:	0
US-10-054-935-2 (1-614) x CN355142 (1-570)			
Oy	310	GluthSerGlnThrLeuProProlyseProPheSerCysGlyArgSerGlyLyseGlyHis	329
Db	2	GAGCATTCCAGATCTGCTCTCCCAACCCCTTCTCATGTGGCGGAAGTGAGAAAGGACAT	61
Oy	330	LysArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLeuAlaPro	349
Db	62	AAAGGAATCCCATTTGGAAAGTACAGAAAGAAAGAACTCTGTTTAAAGCTGCTCCT	121
Oy	350	GluPheSerLysValLysThrLysThrProLysHisSerProIleLysGluGluProCys	369
Db	122	GAATTTTCAAAAGTCAAAACAAAACCTTAAAGACCTCTTATTTAAAGAGAAACCTGT	181
Oy	370	GlySerLeuSerGluThrValCysLysArgGluLeuArgSerGlnGluThrProGluLys	389
Db	182	GGTTCCTTATCTGAAACTGTTTGTGTAACGTGAATTAGAGACCAAGAACCCAGAAAG	241
Oy	390	ProArgSerSerValaAspThrProProArgLysSerThrProGlnLysGlyProSerThr	409
Db	242	CCCCGGCTTCAGAGACACCCCAACCAAGCTCCACCTCCCAAAAAGGAGCCAGACAC	301
Oy	410	HisProLysGluLysAlaPheSerSerGluIleGluAspLeuProLysLysThrThr	429
Db	302	CATCCCAAGAGAAAGGCTTCTCAAGAGATGAAATTAAGATTTGGCCGTTCACACCA	361
Oy	430	GluMetThrLysLeuCysArgTyrHisGlnProProProSerProLeuProLeuArgGluSer	449
Db	362	GAATGTATTTGTGTGCTTGACACCAAGCTCCCATCACCGTTACATTACGGGAATCC	421
Oy	450	SerProLysLysGluGluThrValAla	458
Db	422	TCTCCAAAGAGAGAGACTGTAGCA	448
RESULT 33			
LOCUS	BIS17294	841 bp	mRNA linear EST 29-AUG-2001
DEFINITION	603041522F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5163004 5', mRNA sequence.		
ACCESSION	BIS17294		
VERSION	BIS17294.1	GI:15342086	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.		
TITLE	1 (bases 1 to 841)		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cga@db-jrcmail.nih.gov		
	Tissue Procurement: Life Technologies, Inc.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	Plate: LLM11404 row: k column: 05		
	High quality sequence stop: 815.		

FEATURES
source

Location/Qualifiers
1. 841
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5163004"
/lab_host="DH108"
/clone_1db="NIH MGC 116"
/note="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC library."

ORIGIN

Alignment Scores:
Pred. No.: 3.38e-129 Length: 841
Score: 148.00 Matches: 148
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.10% Indels: 0
DB: 4 Gaps: 0

US-10-054-935-2 (1-614) x B1517294 (1-841)

QY 311 ThrsrGlnThrleuProPolysProPheSerCysGlyArgSerGlyLysGlyHisLys 330
DB 2 ACATCCAGACTCTGCTCCCAAGCCCTTCTCATGTGGCCGAGTGGAAAGGACATAAA 61
QY 331 ArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLysLysLys 350
DB 62 AGGAATATCCCATTTGGAGTACAGAAAGAAAGCTCTGTTAAAGTGGCTCCGAA 121
QY 351 PheSerLysValLysThrLysThrProLysHisSerProLysLysLysLysLysLys 370
DB 122 TTTTCAAAAGTCAAAGAAAGAAAGCTCTCAAGCACTCTCTATTAAGAGAAAGCTGTGCT 181
QY 371 SerLeuSerGluThrValCysLysArgLysLysLysLysLysLysLysLysLys 390
DB 182 TCTTATCTGAAGCTGTTTGTAAAGTGAATTCAGAGCCCAAGAAAGCCCAAGAGGCC 241
QY 391 ArgSerSerValAspThrProArgLysSerThrProGlnLysGlyProSerThrHis 410
DB 242 CGGCTTCAGTGGACACCCCAAGAGCTCTCCCACTCCCAAGAGAGGAGCCAGACCAT 301
QY 411 ProLysGluLysAlaPheSerSerGluLysLysLysLysLysLysLysLysLys 430
DB 302 CCCAAGAGAGAAAGCTCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 361
QY 431 MetThrLeuCysArgGlyThrHisGlnProProSerProSerProLeuArgLysSerSer 450
DB 362 ATGATTGTCGTGGACACAGCTCTCCCATCAGCGTAAACATTAACGGAATCTCTCT 421
QY 451 ProLysLysGluGluThrValAla 458
DB 422 CCAAG 445

RESULT 34
LOCUS BF907636 501 bp mRNA linear EST 18-JAN-2001
DEFINITION QY1-UT0092-041000-411-a06 UT0092 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF907636
VERSION BF907636.1 GI:12299094
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 501)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,M.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.P., Matsushima,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jorgensen,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

PUBMED

10737800

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV1a2=QV1-UT0092-
041000-411-a06&t3=2000-10-04&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 20
High quality sequence stop: 501.

FEATURES
source

Location/Qualifiers
1. 501
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_1db="UT0092"
/note="Organ: uterus tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from OESTRES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Alignment Scores:
Pred. No.: 1.61e-127 Length: 501
Score: 146.00 Matches: 146
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.78% Indels: 0
DB: 2 Gaps: 0

US-10-054-935-2 (1-614) x BF907636 (1-501)

QY 232 Leu1LeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 251
DB 52 CTCATCGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 111
QY 252 SerGluArgAspThrLeuLeuAlaArgLysLysLysLysLysLysLysLysLys 271
DB 112 TCAGAGAGAGACACGCTCTTGCCCGGATTGAAGCATGGAAGAGCGGATGACGTGTA 171
QY 272 LysLysAspAsnGlnLysGluArgHisLysLeuPheGlnGlnGlnGlnGlnGlnGln 291
DB 172 AAGAGAGATTAACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 231
QY 292 GluGluThrGluLeuSerGluLysLysLysLysLysLysLysLysLysLysLysLys 311
DB 232 GAGGAACACAGCTATCTGAGAAATTAACATGGAAGGCCAGCGAGCTTCCGAGACA 291
QY 312 SerGlnThrLeuProPolysProPheSerCysGlyArgSerGlyLysGlyHisLysArg 331
DB 292 TCCAGACTCTGCTCCCAAGCCCTTCTCATGTGGCGGAGTGAAGAGGAGACATAAAGG 351

QY 332 lvsSerProheglsrThrgluArgLysThrProValylsLeuAlaProgluPhe 351
 |||||
 Db 352 AAATCCCATTTTGAAGTACGAAAGAACCTCTGTATAAAGCGCTCCGATTT 411

QY 352 SetlValylsThrlyrThrProLysHisSerProIlelsgluGluProCysGlySer 371
 |||||
 Db 412 TCAAAGTCAAAACAAACAACTCTTACGACTCTCTATTAAAGGAAACCTGTGTTCC 471

QY 372 leuSerGluThrValCys 377
 |||||
 Db 472 TTATCTGAACCTGTTGT 489

RESULT 35
 AV762388 699 bp mRNA linear EST 19-OCT-2000
 AV762388 MDS Homo sapiens cDNA clone MDSAJG10 5', mRNA sequence.
 AV762388
 AV762388
 AV762388.1 GI:10920236
 EST.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 699)
 Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H.,
 Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
 Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M.,
 Lu,G., Yang,Y., Gao,G., Zhang,Q., Chen,S., Han,Z. and Chen,Z.
 Homo sapiens cDNA MDS clones
 Unpublished (2000)

TITLE
 JOURNAL
 COMMENT
 Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919 (ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 Location/Qualifiers
 1..699
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MDSAJG10"
 /tissue_type="Bone marrow"
 /cell_type="CD34+ hematopoietic stem/progenitor cell"
 /lab_host="BM25.8"
 /clone_lib="MDS"
 /note="Vector: pTriplEx2; Site_1: sflta; Site_2: sfltb"

ORIGIN

Alignment Scores:
 Pred. No.: 1.75e-125 Length: 699
 Score: 144.00 Matches: 144
 Percent Similarity: 100.00% Conservat: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 23.45% Indels: 0
 DB: 1 Gaps: 0

US-10-054-935-2 (1-614) x AV762388 (1-699)

QY 296 leuSerGluThrlyrThrProLysHisSerProIlelsgluGluProCysGlySer 315
 |||||
 Db 21 CTATCTGAGAAATTAACTGAGTGCACGCCGAGTTTCCGAGACATCCGACCTG 80

QY 316 ProProLysProPheSerCysGlyArgSerGlyLysGlyHisLysAlaGlySerProPhe 335
 |||||
 Db 81 CCTCCCAAGCCCTTCATCATGTGGCGGAGTGAAGGACATATAAAGAAATCCCATTT 140

QY 336 GlySerThrGluArgLysThrProValylsLeuAlaProgluPheSerlyValyls 355
 |||||
 Db 141 GGAAGTACAGAAAGAAAGACTCTGTATAAAGCTGCTCTGTAATTTTCAAAAGTCAA 200

QY 356 ThrlyrThrProLysHisSerProIlelsgluGluProCysGlySerleuSerGluThr 375
 |||||
 Db 201 ACAAAAGCTCTAAGACACTCTCTATTAAAGGAACCTGTGTCTTATCTGAACCT 260

QY 376 ValCysAlaArgGluLeuArgSerGluGluThrProGluLysProArgSerValArg 395
 |||||
 Db 261 GTTTGTAACGTAATTGAGAGACCAAGAAACCCGAGAAAGCCCGGTCTTCATGTGAC 320

QY 396 ThrProArgLeuSerThrProGluLysGlyProSerThrHisProLysGluValAla 415
 |||||
 Db 321 ACCCCACCAAGACTCTCCACTCCCAAAAGGACCAAGACCATCCCAAGAGAAAGCC 380

QY 416 PheSerSerGluIleGluAspLeuProTyrLeuSerThrGluMetTyrLeuValArg 435
 |||||
 Db 381 TTCTCAAGTACAGATAGAAATTTGCGTACTCTTCACACAGAAATGTATTGTGCTGT 440

QY 436 TrpHisGluPro 439
 |||||
 Db 441 TGGACACCACT 452

RESULT 36
 BM728253 428 bp mRNA linear EST 01-MAR-2002
 LOCUS
 DEFINITION
 UI-E-EJ0-ain-h-18-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
 UI-E-EJ0-ain-h-18-0-UI 5', mRNA sequence.
 ACCESSION
 BM728253
 BM728253.1 GI:19049586
 EST.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 428)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtracction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 8889548

JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 WEBER, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..428
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-EJ0-ain-h-18-0-UI"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina foveal and Macular, RPE and
 Choroid"
 /dev stage="fetal and adult"
 /lab_host="DH10B (life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-EJ0"
 /note="Organ: eye; Vector: pTripl-Pac (pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-EJ0 is a subtracted cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested

ORIGIN

Alignment Scores:

Pred. No.:	8.53e-124	Length:	428
Score:	142.00	Matches:	142
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	23.13%	Indels:	0
DB:	4	Gaps:	0

US-10-054-935-2 (1-614) x BW728253 (1-428)

with Not 1, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (drr)18 tail. The sequence tags for this library are: fetal eyes, AGATCAGAG, lens, CCATTAGCA, eye anterior segment, AATGCCGCTAT, optic nerve, CCATTAGCT, retina, CCGCG, Retina Foveal and Macular, GTCC, RPE and Choroid, ACCCA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

```
QY 210 G1yG1yG1ySerG1yAlaSerSerG1nAla1aCysleuylsg1n1leleu1euln 225
Db 1 GGTGGCGGCTCGGAGGCTCCAGTCAGGCGCGCTGCTCAAAAGATCTTCTGCTCAA 60
QY 230 LeuAspLeu1leG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1n 249
Db 61 TTGGACCTCTCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120
QY 250 Leu1ySerG1yArG1yArG1yArG1yArG1yArG1yArG1yArG1yArG1yArG1y 269
Db 121 CTGAAGTCAGAGAGAGACAGCTCTTGTGATGATGATGATGATGATGATGATGATG 180
QY 270 LeuVal1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y 289
Db 181 CTGGTAAAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 290 G1uArgG1uG1uThrG1uLeuSerG1u1y1y1y1y1y1y1y1y1y1y1y1y1y 309
Db 241 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 310 G1uThrSerG1nThrLeuProProProProProProProProProProProPro 329
Db 301 GAGGATCCAGACTGCTGCTCCCAAGCCTTCTCAATGGGCGAGAGAGAGAGAG 360
QY 330 1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y 349
Db 361 AAAAGAAATCCCATTTGGAGATACAGAAAGAAAGAAAGAAAGAAAGCTGCTCT 420
QY 350 GluPhe 351
Db 421 GAATTT 426
RESULT 37
AM192854 463 bp mRNA linear EST 29-NOV-1999
LOCUS x154d12.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:678519 3'
DEFINITION similar to contains_PIR5.t3 TAR1 repetitive element ;, mRNA
sequence.
ACCESSION AM192854
VERSION AM192854.1 GI:6471553
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 463)
AUTHORS NCI-CGAP http://www.nci.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
```

FEATURES

source

Email: cgaps-remail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdrp/image/image.html
Seq primer: -400P from G1bco
High quality sequence stop: 441.
Location/Qualifiers
1. 463
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2678519"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_1lb="NCI_CGAP_Pan1"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

ORIGIN

Alignment Scores:

Pred. No.:	9.21e-124	Length:	463
Score:	142.00	Matches:	142
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	23.13%	Indels:	0
DB:	2	Gaps:	0

US-10-054-935-2 (1-614) x AM192854 (1-463)

```
QY 10 A1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 29
Db 2 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61
QY 30 A1a1eul1yG1yProG1yAspG1yProG1yAla1a1a1a1a1a1a1a1a1a1a 49
Db 62 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 121
QY 50 Arg1y1eul1yG1yProG1yProProProProProProProProProProProPro 69
Db 122 CGTAAGCTCAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 181
QY 70 SerPro1a1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y 89
Db 182 TCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241
QY 90 G1y1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1n 109
Db 242 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
QY 110 1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y 129
Db 302 AAGCAAGCGGCAATTTGGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
QY 130 1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y1y 149
Db 362 AAGTATCAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421
QY 150 ThrPro 151
Db 422 ACGGCC 427
RESULT 38
BF116257 479 bp mRNA linear EST 24-OCT-2000
LOCUS BF116257
DEFINITION 7a79n02.x1 NCI CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3571035 3'
sequence.
ACCESSION BF116257
VERSION BF116257.1 GI:10985733
```

	KEYWORDS	EST.
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 479)	
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapb-remail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LMUL, send email to: info@image.lmnl.gov Seq primer: -40UP from Glbpco High quality sequence stop: 407. Location/Qualifiers	
FEATURES	source	
	1..479	
	/organism="Homo sapiens"	
	/mol_type="rRNA"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:3571035"	
	/tissue_type="fibrotheoma"	
	/lab_host="DH10B (phage-resistant)"	
	/clone_lib="NCI_CGAP_Ovi8"	
	/note="Organ: ovary; Vector: p7TJD-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTCGGAGCGCGCCGCAGCATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia). digested with Not I and cloned into the Not I and Eco RI sites of the modified p7TJ3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."	
ORIGIN		
Alignment Scores:		
Pred. No.:	9.52e-124	Length: 479
Score:	142.00	Matches: 142
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	23.13%	Indels: 0
DB:	2	Gaps: 0
US-10-054-935-2 (1-614) x BF116257 (1-479)		
Oy	10	AlAa1aa1aa1aPro1aAGlyYAsnpProgiugniArgyleuaApTYrrgLuAa1aala 29
Dd	1	GGGGCCCCGCCCTGTGCGGGCAGATCTGAAGCAGCATGAGTAAGAAGCGGCTCG 60
Oy	30	AlAleugiLygiProGiLuAsPgluProgiYalAa1agiua1ahisPheLeuProARhis 49
Dd	61	GGCGTGGGGCGGGCCGAGAGCAGAGCTGTGGGGCGCGCGAAGCCCACTTCTCCCCGGGAC 120
Oy	50	ArglysleuLygsLupProgiYlProProLeua1aserSerGingLygiSerProAlaPro 69
Dd	121	CGTAAAGCTCAAAGAGCCGGGGCCCCCGCTGGCTCTCTCCAGAGGGGGAGAGCCCGCGCT 180
Oy	70	SerPro1aagiCyAgLygiLygiLygiLygiLeuLeuLeuPro1aagiYAla1aPro 89
Dd	181	TCCCCGGGGCTGGGGCGGAGAGGCGCGGGCTTTGTAATCCGGCGGGGGCGCCCCC 240
Oy	90	GlyginginigiugLiuserTrigLygiSerValProLeuProCyPProProAlaThr 109
Dd	241	GGGACAGAGAGAGACTGGGGCGGCTTCGATGCTCCCTTGGCTTCGCGCCCCCGCACCC 300
Oy	110	LysGlna1agiLygieLygiLygiLupPro1a1a1aagiYagiCySerProArGPro 129

Db	301	AAGCAACCGGCATTTGGGGGGGAGCGCTGCCGACCGCGACCGCGCTGCAGCCCCCGGCC	360
Qy	130	LyfYtGlnAlaValLeuP ¹ collglnThrGlySerleuValAlaAlaAlaValGluPro	149
Db	361	AAGATCATGGCGGGTGGTCCCATTCAGACGGGCTCTCTCGTGGCGGGCGCAAGAGCTT	420
Qy	150	ThrPro 151	
Db	421	ACGCC 426	
RESULT 39			
LOCUS	BI667371	551 bp	mRNA linear EST 12-SEP-2001
DEFINITION	60329214F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5311571 5',		
ACCESSION	BI667371		
VERSION	BI667371.1		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 551)		
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
CONTACT	Contact: Robert Strausberg, Ph.D.		
EMAIL	Email: cgabs@emall.nih.gov		
TISSUE	Tissue Procurement: Miklos Palkeovics, M.D., Ph.D.		
CDNA	CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki		
Toshiyuki and Piero Carninci (RIKEN)			
CDNA	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
DNA	DNA sequencing by: Incyte Genomics, Inc.		
CLONE	Clone distribution: MGC clone distribution information can be		
FOUND	found through the I.M.A.G.E. Consortium/LNL at:		
HTTP	http://image.lnl.gov		
PLATE	Plate: LHM11789 row: 1 column: 12		
HIGH	High quality sequence stop: 551.		
LOCATION	Location/Qualifiers		
1..551			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:5311571"			
/tissue_type="hypothalamus"			
/lab_host="DH10B"			
/clone_lib="NIH_MGC_96"			
/note="Organ: brain; Vector: BluescriptPR (modified			
pbuescript KS+); Site_1: BamI; Site_2: SalI-XhoI			
(gtagcag); Oligo-dT primed using primer			
5'-TTTTTTTTTTTTTAA-3', size-selected for average			
insert size 2.3 kb and normalized to R07.5. This is a			
primary library enriched for full-length clones and			
constructed using the Cap-trapper method (carninci, in			
preparation). Library constructed by M. Brownstein			
(NIH/NHGRI, National Institutes of Health). Note: this is			
a NIH_MGC library."			
ORIGIN			
Alignment Scores:			
Prod. No.:	7.6e-121	Length:	551
Score:	139.00	Matches:	182
Percent Similarity:	98.91%	Conservative:	0
Best Local Similarity:	98.91%	Mismatches:	1
Query Match:	22.64%	Indels:	2
DB:	4	Gaps:	0
US-10-054-935-2 (1-614) x BI667371 (1-551)			
Qy	263	ArgMetGlnArgMetGlnLeuValIyLeuAspAsnGluIyGlnArgHisIyLeu	282
Db	3	CGGATGGAAGGCGGATGACGTGGTAAAGATTAACGAGAAAGAAAGCACAAAGCTG	62

QY 283 PheGlnGlyTyrGlnThrGlnGluGluGluThrGluLeuSerGluIleuValLeu 302
Db 63 TTCAGGGCTATGAACTGAAAGAGAGAGAAACAGAGCTATCTGAGAAATTAAC 122
QY 303 GluCyseGlnPro-GluLeuSerGluThrSerGlnThrLeuProProIleProPheSerCy 322
Db 123 GAGGCGCAGCGG-GGAGGCTTTCGAGACATCCACACTCTGCTCCCAAGCCCTTCATG 181
QY 322 sGlyArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGlnArgLys 342
Db 182 TGGGCGGAGTGGAAAGGACATTAAGAAATCCCATTTGGAGTACAGAAAGAAAGAC 241
QY 342 rProValIleLysLeuAlaProGluPheSerLysValIleThrLysThrProLysHis 362
Db 242 TCCGTTAAAGAGCTGCTCTCGAATTTTCAAAAGTCAAAACAAAACCTCTTAAGCATC 301
QY 362 rProIleLysGluGluProCyseGlySerLeuSerGluThrValLysLysArgGluLeu 382
Db 302 TCCATTAAAGAGAACCTGTGCTCTTATCTGAAACGTGTTGTAACGTGAATGAG 361
QY 382 gSerGlnGluThrProGluLysProArgSerSerValAspThrProProArgLeuSer 402
Db 362 GAGCCAGAAACCCCAAGAAAGCCCGGCTCTTCACTGAGACCCACCAAGACTCTCAC 421
QY 402 rProGlnLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIleGlu 422
Db 422 TCCCAAAAGAGAACCCAGACCCATCCCAAGAGAAAGCCCTTCACTGAGATGAGA 481
QY 422 pLeuProTyrLeuSerThrThrGluMetTyrLeuCyseArgTyrHisGlnProPro 442
Db 482 TTGGCCCTACTTCCACACAGAAATGATTTGTGTGCTGGCAGCAGCCCTCCCATC 541
QY 442 rProLeuPro 445
Db 542 ACCGTACCA 551

RESULT 40
LOCUS BM475963 1032 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6480858 NIH_MGC_67 Homo sapiens CDNA clone IMAGE:5579934
ACCESSION BM475963
VERSION BM475963.1 GI:18525014
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1032)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1AM1237 row: 9 column: 07
High quality sequence stop: 653.
Location/Qualifiers
1. 1032

FEATURES
Source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5579934"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_67"

/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

ORIGIN

Alignment Scores:

Pred. No.:	1,4e-120	Length:	1032
Score:	139.00	Matches:	139
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	22.64%	Indels:	0
DB:	4	Gaps:	0

US-10-054-935-2 (1-614) x BM475963 (1-1032)

QY 476 ProSerTTPArgAspHisSerValGluProLeuArgAspProAsnProSerAspLeu 495
Db 44 CTTCTTGGAGGAGCACCTCAGTAGAGCTCTTAAGGAGCAACCAATCTTCAGACTTTTG 103
QY 496 GluAsnLeuAspAspSerValPheSerLysArgHisAlaLysLeuGluLeuAspGluLys 515
Db 104 GAGAACCTTGATGACAGCTGTGTTTTCAGACCGGATGCAAACTGAGCTGATGAGAG 163
QY 516 ArgArgLysArgTTPAspIleGlnArgIleArgGluGlnArgIleGlnArgLeuGln 535
Db 164 AGAAGGAAAGATGGGATATTGAGAGATCAGGAGCAAGAAATTTTACAGCGACTG 223
QY 536 LeuArgMetTyrLysLysLysGlyIleGlnGluSerGluProGluValThrSerPhe 555
Db 224 CTCAGATGTATTAAGAAAGAAAGAAATTCAGAAATTCGAGCTGAGTACCTCATTTTTC 283
QY 556 ProGluProAspAspValGluSerLeuMetIleThrProPheLeuProValAlaPhe 575
Db 284 CTTGAGCCAGATATGTGAAGTTGATGATTAACCCCTTCTTGTGCTGTGATGACATTT 343
QY 576 GlyArgProLeuProLysLeuThrProGlnAsnPheGluLeuProTyrLeuAspGluArg 595
Db 344 GAGACGACATTAACCAAAATTAATCTCCACAGAAATTTGAGTACCTGTTGATGAGCGT 403
QY 596 SerArgCyseArgLeuGluIleGlnLysLysGlnThrProHisArgThrCyseArgLys 614
Db 404 AGCCGATGACAGATGAGATCCAGAAAGCAAACTCACCGGACGTGATGAGAA 460

RESULT 41
LOCUS BG421221 1087 bp mRNA linear EST 14-MAR-2001
DEFINITION 602451712F1 NIH_MGC_14 Homo sapiens CDNA clone IMAGE:459002 5',
mRNA sequence.
ACCESSION BG421221
VERSION BG421221.1 GI:13327727
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1087)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1CM1324 row: h column: 03
High quality sequence stop: 583.
Location/Qualifiers
1. 1087

FEATURES
Source

RESULT 43
 BF515361
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 SOURCE

BF515361 524 bp mRNA linear EST 07-DEC-2000
 UI-H-BM1-ann-c-03-0-UI.s1 NCI_CGAP_Sub7 Homo sapiens cDNA clone
 IMAGE:3082733 3', mRNA sequence.
 BF515361 GI:11600540
 BF515361
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 524)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 oligo-dt track not found. Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.B.
 Soares Lab Clone distribution: NCI-CGAP clone distribution
 information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrrp/image/image.html
 Seq primer: M13 Forward
 POLYA=NO.

Location/Qualifiers
 1..524
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3082733"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI_CGAP_Sub7"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not 1; Site 2: Eco RI; NCI CGAP Sub7
 is a subtracted library derived from NCI CGAP Sub6. The
 NCI CGAP Sub7 library had 12 million recombinants. A
 single-stranded DNA preparation of NCI CGAP Sub6 was used
 as a tracer in a subtractive hybridization with a driver
 comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM
 3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids
 132376-132391, 1456008-1456775, 1500552-1502855);
 NCI CGAP Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
 (IMAGE Clonoids 1323912-1325831, 1471368-1472903,
 1492104-1493255); NCI CGAP L45 pool 1 LLAM 3575-3582,
 3851-3854 (IMAGE Clonoids 1414920-1417991,
 1520904-1522439); NCI CGAP GC4 pool 1 LLAM 3164-3167,
 3716-3720, 3733-3735 (IMAGE Clonoids
 1257096-1258631, 1469064-1470983, 1475592-1476743);
 NCI CGAP Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
 (IMAGE Clonoids 985608-986759, 1101192-1101959,
 1217928-1220615); NCI CGAP Co10 pool 1 LLAM 2644-2653,
 2871-2872 (IMAGE Clonoids
 1057416-1061255, 1144584-1145351). (6% of the driver
 population), plus a pool of 3,840 arrayed clones from
 NCI CGAP Sub1 (IMAGE Clonoids 2708616-2710535) and
 NCI CGAP Sub2 (IMAGE Clonoids 2710536-2712455) (4% of
 the driver population), plus a pool of 11,136 clones from
 NCI CGAP Sub3 (IMAGE Clonoids 2712456-2723591) (10% of
 the driver population), plus a pool of 5,472 clones from
 NCI CGAP Sub4 (IMAGE Clonoids 2723592-2729336) (40% of the
 driver population), plus a pool of 4032 clones from
 NCI CGAP Sub6 (IMAGE Clonoids 2728969-2733190) (40% of the
 driver population). Subtraction was performed as
 previously described (Bonaldo, Lennon & Soares (1996) :
 Normalization and Subtraction: Two Approaches To
 Facilitate Gene Discovery. Genome Research 6, 791-806.
 TAG_SEO=None found"

Pred. No.: 3,976-116 Length: 524
 Score: 134.00 Matches: 134
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 21.82% Indels: 0
 DB: 2 Gaps: 0

US-10-054-935-2 (1-614) x BF515361 (1-524)

QY 225 |||leu|leu|leu|gln|leu|asp|leu|ile|gln|gln|gln|gln|leu|gln|ala|alys|glu 244
 |||
 Db 121 ATCTTGTGCTGCAATTTGACCTTCATCGAACAGACGACGACGCTTCCAGCCAAAGGAA 180
 QY 245 lye|glu|ile|glu|leu|leu|lys|ser|glu|arg|asp|thr|leu|ala|arg|ile|glu|arg|met 264
 |||
 Db 181 AAGGAGATCGACGAGAGCTGAAGTCAAGAGAGACACGCTCTTGTGCGATTGAACGATG 240
 QY 265 glu|arg|arg|met|gln|leu|val|lys|lys|asp|asn|glu|lys|glu|arg|his|lys|leu|phe|gln 284
 |||
 Db 241 GAAAGCGGATGACGCTGTAAAGAGATTAACGAGAAAGGACCAAGCTGTTTCAG 300
 QY 285 gl|tyr|glu|thr|gln|glu|arg|glu|gln|thr|gln|leu|ser|glu|val|ile|lys|leu|gln|cys 304
 |||
 Db 301 GCGTATGAAGCTAAGAGAGAGAGAGAAACAGACTATCTAGAGAAATTAACCTGAGTGC 360
 QY 305 gl|pro|glu|leu|ser|glu|thr|ser|gln|thr|leu|pro|pro|lys|pro|phe|ser|cys|glu|arg 324
 |||
 Db 361 CAGCCGAGCTTTCCGAGACATCCGACATCTGCTCCAGCCCTTCTCATGTGGGCGG 420
 QY 325 ser|glu|lys|glu|his|lys|arg|lys|ser|pro|phe|glu|ser|thr|glu|arg|lys|thr|pro|val 344
 |||
 Db 421 AGTGAAGAGGACATTAAGAAATATCCCATTTGGAAGTACAGAAAGAAAGAACTCTGTT 480
 QY 345 lys|lys|leu|ala|pro|glu|phe|ser|lys|val|lys|thr|thr 358
 |||
 Db 481 AAAAAGCTGGCTCCTGTAATTTTCAAAAGTCAAACAAAAACT 522

RESULT 44
 CB268936
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

CB268936 580 bp mRNA linear EST 20-FEB-2003
 1007843 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
 cDNA 5', mRNA sequence.
 CB268936
 CB268936.1 GI:28443521
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 580)
 Yang, R.-Z., Shuldiner, A. and Gong, D.-W.
 ESP analysis of human adipose gene expression
 Unpublished (2002)
 Contact: Gong Da-Wei
 Division of Endocrinology, Diabetes and Nutrition
 University of Maryland
 660 Redwood St, H497, Baltimore, MD 21201, USA
 Tel: 410 706 1622
 Fax: 410 706 1622
 Email: dgong@medicine.umaryland.edu
 PCR Primers
 FORWARD: CTGGGAAGCGCCCATTTGCTGTG
 BACKWARD: AATAGCATCATATTAAGGCGCAATTGG
 Seq primer: GTTGGTACCGGGAATTC.
 Location/Qualifiers

FEATURES
 SOURCE
 1..580
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /sex="Male and Female"
 /tissue_type="Adipose"
 /clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
 /note="Vector: lambda triplex"

Alignment Scores:

ORIGIN

Alignment Scores:

Pred. No.: 3.89e-115 Length: 580

Score: 133.00 Matches: 133

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 21.66% Indels: 0

DB: 6 Gaps: 0

US-10-054-935-2 (1-614) x CB268936 (1-580)

QY 327 LysG[His]LysSerProPheG[Ser]ThrGluArgLysThrProValLys 346

DB 1 AAGGGAATTAAGAAATCCCTTTGGAAGTACAGAAAGAAAGACTCTGTTAAAG 60

QY 347 LeuAlaProGluPheSerLysValLysThrLysThrProLysHisSerProIleLysGlu 366

DB 61 CTGGCTCTGAATTTTCAAAAGTCMAAACAAAACCTTAAGCACTCTCTATTAAAG 120

QY 367 GluProCysGlySerLeuSerGluThrValCysLysArgLysLeuAArgSerGlnGluThr 386

DB 121 GAACCTGTGTCTTATCTGAAGCTGTTGTAAGTGAAGAGCCAGAAAGACC 180

QY 387 ProGluLysProArgSerSerValAspThrProProArgLeuSerThrProGlnLysGly 406

DB 181 CCAAGAAAGCCCGGTCTTCAAGTGAACACCCCAAGACTCTCCACCAAAAGGA 240

QY 407 ProSerThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuProLysLeu 426

DB 241 CCAGACACCATCCCAAGAGAAAGCTTCTCAAGTGAAGTGAAGTTGGCCGTACTT 300

QY 427 SerThrThrGluMetLysLysCysArgTyrPheGlnProProSerProLeuProLeu 446

DB 301 TCCACACAGAAATGATTTGTGCGTGGACACAGCCCTCCCATGACCGTAACTTA 360

QY 447 ArgGluSerSerProLysLysGluGluThrValAlaArg 459

DB 361 CGGGAATCTCTCCAAAGAGAGAGACTGTAGCAAG 399

RESULT 45

AM965141 686 bp mRNA linear EST 01-JUN-2000

LOCUS EST377214 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.

DEFINITION AM965141

ACCESSION AM965141 GI:8154977

VERSION AM965141.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 686)

AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.

TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

JOURNAL Unpublished (2000)

COMMENT The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528

Fax: 301 838 0208

Email: johnq@tigr.org

Plate: 210

Seq primer: Reverse.

FEATURES

source

1..686

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGI"

/note="Vector: pBluescriptSkm"

ORIGIN

Alignment Scores:

Pred. No.: 1.97e-108 Length: 686

Score: 126.00 Matches: 192

Percent Similarity: 98.97% Conservative: 0

Best Local Similarity: 98.97% Mismatches: 1

Query Match: 20.52% Indels: 2

DB: 2 Gaps: 0

US-10-054-935-2 (1-614) x AM965141 (1-686)

QY 229 GlnLeuAspLeuIleGluGlnGlnGlnGlnLeuGlnAlaLysGluLysGluIleGlu 248

DB 3 CAATTGAGCTCTATCCAGACGACGACGACCTCAGGCCAAGAGAAAGAGATTCGAG 62

QY 249 GluLeuLysSerGluArgAspThrLeuLeuAlaArgIleGluArgMetGluArgMet 268

DB 63 GACCTGAAGTCAGAGAGACGACGCTCTTGTCTCGATTGAACGTATGAAAGCCGATG 122

QY 269 GlnLeuValLysLysAspAsnGluLysGluArgHisLysLeuPheGlnGlyTyrGluThr 288

DB 123 CAGCTGTTAAAGAAAGATTAACGAAAGAAAGGCAAGCTGTTTCAGGGCTATGAAGT 182

QY 289 GluGluArgGluGluThrGluLeuSerGluLysIleLysLeuGluCysGlnProGluLeu 308

DB 183 GAAGAGAGAGAGAAACAGATCTTATCTGAAGAAATTAACCTGAGGTGCGAGGAGCT 241

QY 308 uSerGluThrSerGlnThrLeuProProLysProPheSerCysGlyArgSerGlyLysGly 328

DB 242 TTCCGAGACATCCCAAGCTCTGCTCCAGACCTTCTCATGTGGCGAGTGGAAAGG 301

QY 328 YHisLysArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLeuAl 348

DB 302 ACATAAAGAAATCCCAATTTGGAAGTACAGAAAGAAAGACTCTGTTAAAAAGCTGGC 361

QY 348 aProGluPheSerLysValLysThrLysThrProLysHisSerProIleLysGluLysP 368

DB 362 TCTGGAATTTTCAAAAGTCMAAACAAACTCTTAAGCACTCTCCATTAAGAGAAAG 421

QY 368 oCysGlySerLeuSerGluThrValCysLysArgLysLeuArgSerGlnGluThrProG 388

DB 422 CTGTGTTCTTATCTGAAGCTGTTGTAAACGTGAATGAGAGCCMAAGAAACCCAGA 481

QY 388 uLysProArgSerSerValAspThrProProArgLeuSerThrProGlnLysGlyProSe 408

DB 482 AAAGCCCGGTCTTCAAGTGAACCCCAAGACCTCCACCTCCCAAAAGGAGCCAG 541

QY 408 rThrHisProLysGluLysAlaPheSerSerGluIleGlu 421

DB 542 CACCCATCCCAAGAGAAAGCTTCTCAAGTGAAGATGAA 581

RESULT 46

AA650422 442 bp mRNA linear EST 13-NOV-1997

LOCUS ns95h06.s1 NCI_CGAP_P33 Homo sapiens cDNA clone IMAGE:1191419, mRNA sequence.

DEFINITION AA650422

ACCESSION AA650422

VERSION AA650422.1 GI:2577750

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 442)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncigap>.

TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528

Fax: 301 838 0208

Email: cga@nci.nih.gov

Plate: 210

Seq primer: Forward.

FEATURES

source

1..442

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone_lib="NCI-CGAP Tumor Gene Index"

/note="Vector: pBluescriptSkm"

CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.linn.gov/bdrp/image/image.html
Insert Length: 759 Std Error: 0.00
Seq primer: -40m13 fwd. ER from Amersham
High quality sequence stop: 411.
Location/Qualifiers

FEATURES

1..442
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1191419"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/clone_lib="NCI CGAP Pr3"
/note="Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected cells
histologically-determined to be fully malignant prostate
cancer cells. Double-stranded cDNA was ligated to EcoRI
adaptors, 5 cycles of PCR applied to the cDNA with an
adaptor-specific primer, and the resulting PCR product
subcloned into pAMP10 by the UDG-cloning method (Life
Technologies). Average insert size is 600 bp. NOTE: Not
directionally cloned. This library was constructed by
David Krizman."

ORIGIN

Alignment Scores:

Pred. No.: 1.14e-107 Length: 442
Score: 125.00 Matches: 125
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.36% Indels: 0
Gaps: 0

US-10-054-935-2 (1-614) x AA650422 (1-442)

QY 323 G1AArgSerGlyLysGlyHisLysArgLysSerProPhagGlySerThrGluArgLysThr 342
DB 10 GGGGGAGTGGAAAGGACATTAAGAAATCCCATTTGAAATACGAAAGAAAGACT 69
QY 343 ProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysHisSer 362
DB 70 CCGGTAAAGAGTGGCTCCGTAATTTCAAAAGTCAAACAAACCTCTTACACACTCT 129
QY 363 ProLeuLysGluGluProCysGlySerLeuSerGluThrValCysLysArgLysLeuArg 382
DB 130 CCAATTAAAGAGAACCTGCTGTTCTTAACTGAAACCTGTTTAAACGGAATTGAGG 189
QY 383 SerGlnGluThrProGluLysProArgSerSerValAspThrProProArgLysSerThr 402
DB 190 AGCCAAAGAAACCCAGAAAGCCCCGGCTTCAATGACACCCCAACAGACTCTCACT 249
QY 403 ProGlnLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluLysLeuAsp 422
DB 250 CCCCAGAAAGGAGCCACACACCATCCCAAGAGAAAGCCTTCTCAAGTGAATGAAGAT 309
QY 423 LeuProLysLeuSerThrThrGluMetLysCysArgTgPheGlnProProProSer 442
DB 310 TTGCGGACCTTTCACACAGAAATGATTTGTGTGGACCAAGCTCCCTCCCATCA 369
QY 443 ProLeuProLeuArg 447
DB 370 CCGTTACCATTAACGG 384

RESULT 47
HSM800235

LOCUS HSM800235 1433 bp mRNA linear HTC 22-SEP-2004
DEFINITION Homo sapiens mRNA; cDNA DKFZp586B1922 (from clone DKFZp586B1922).
ACCESSION AL049450
VERSION AL049450.1 GI:4500236
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1433)
Fodor, G., Han, W., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
Koehler, K., Beyer, A., The German CDNA Consortium
Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuerberg, GERMANY
Neuberberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center) at the
Heinrich-Heine-University, Dusseldorf/Germany) within the CDNA
sequencing consortium of the German Genome Project. This clone
(DKFZp586B1922) is available at the RZPD Deutsches
Resequenzierungszentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
<http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp586B1922>
Further information about the clone and the sequencing project is
available at <http://mips.gsf.de/projects/cdna/>.

FEATURES

source

1..1433
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp586B1922"
/db_xref="taxon:9606"
/clone="DKFZp586B1922"
/tissue_type="uterus"
/clone_lib="586 (synonym: hute1). Vector pSport1; host
DH10B; sites NotI + SalI/MluI"
/dev_stage="adult"
/note="unclassified"
1..1433
/gene="DKFZp586B1922"
/c363..>737
/gene="DKFZp586B1922"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH10734.1"
/db_xref="GI:50949271"
/translation="PEGSTERKTPVKKLAFESKVKTKPKSPKPEPGSLSEIVC
KRLRSQETPEKRRSSVDTPRLSTPOKGPSTHPKKAFFSEIEDLPYSTTEWYLCR
WHQPPSPPLPRSSPKKEETVA"

ORIGIN

Alignment Scores:

Pred. No.: 3.58e-107 Length: 1433
Score: 125.00 Matches: 125
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.36% Indels: 0
Gaps: 0

US-10-054-935-2 (1-614) x HSM800235 (1-1433)

QY 334 ProPhagGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSerLys 353
DB 363 CCAATTGAAAGTACAGAAAGAAAGACTCCGTAAAGAGCTGGCTCTGAATTTTCAAA 422
QY 354 ValLysThrLysThrProLysHisSerProLysGluGluLysProCysGlySerLysSer 373
DB 423 GTCAAAACAAAACCTCTTAAGCACTCTTATTAAAGAGAACCTGTGTTCTTATCT 482
QY 374 GluThrValCysLysArgLysLeuArgSerGlnGluThrProGluLysProArgSerSer 393
DB 483 GAAACTGTTGTAAACGTGAATTGAGAGCAAGAAACCCCAAGAAAGCCCCGGCTTTCA 542

QY 394 ValAspThrProProArgLeuSerThrProGlnlyGlyProSerThrHisProlyGln 413
 |||||
 Db 543 GUGGACACCCACCAAGACTCTCCACCCCAAAAGGACCCAGACCCATCCCAAGAG 602
 QY 414 LysAlaPheSerSerGluIleGluAspLeuProTyrLeuSerThrThrimetTyrLeu 433
 |||||
 Db 603 AAGCCTTCCTCAAGTAGATAGAAAGATTTCGCCACTTCCACACAGAAATGTATTG 662
 QY 434 CysArgTTrpHisGlnProProSerProLeuProLeuArgLeuSerProlyGly 453
 |||||
 Db 663 TGTGGTGGACACGACTCCCATCATCATCATCATCATCATCATCATCATCATCAT 722
 QY 454 GluGluThrValAla 458
 |||||
 Db 723 GAGGAGACTGTAGCA 737

RESULT 48
 BG480294 651 bp mRNA linear EST 21-MAR-2001
 LOCUS 602529340F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4653152 5',
 DEFINITION mRNA sequence.
 ACCESSION BG480294
 VERSION BG480294.1 GI:13412573
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE NIH-MGC http://mgs.nci.nih.gov/.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLCM438 row: 0 column: 09
 High quality sequence stop: 651.
 Location/Qualifiers

FEATURES

source

1..651
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4653152"
 /issue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NIH MGC 21"
 /note="Organ: placenta; Vector: pOT87; Site 1: XhoI;
 Site 2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.: 1..31e-105 Length: 651
 Score: 123.00 Matches: 123
 Percent Similarity: 100.00% Conserves: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.03% Indels: 0
 Gaps: 0

US-10-054-935-2 (1-614) x BG480294 (1-651)

QY 231 AspLeuIleGluGlnGlnGlnGlnGlnGlnAlaLeuGluLeuGluIleGluGluLeu 250
 |||||
 Db 1 GACCTATGACACAGACAGACAGCTGACAGCCCAAGAAAGAGATCGAGAGCTG 60
 QY 251 LysSerGlnArgAspThrLeuLeuAlaArgIleGluArgMetGlnArgArgMetGlnLeu 270
 |||||
 Db 61 AAGTCAGAGAGAGACAGCTCTGCTCGGATTTGAACGATGGAAGGGGAGATGACGCTG 120
 QY 271 ValLysLysAspAsnGlnLysGlnLysGlnLysLysLeuPheGlnIleTyrGluThrGluGlu 290
 |||||
 Db 121 GTAAAGAGAGATTAACAGAAAGAAAGCAAGCTTTTCAGGGCTATGAAATCGAAAGAG 180
 QY 291 ArgGluGluThrGluLeuSerGlnLysIleLysLeuGluCysGlnProGluLeuSerGln 310
 |||||
 Db 181 AGAGAGAAACAGAGCTATCTGAAATTAATTAATGAGAGTCCAGCCGAGACTTTCGAG 240
 QY 311 ThrSerGlnThrLeuProProLysProPheSerCysGlyArgSerGlnLysGlnLys 330
 |||||
 Db 241 ACATCCACAGCTCTGCTCCCAAGCCCTTCTCATGTGGCGAGGTGGAAGGACATTA 300
 QY 331 ArgLysSerProPheGlnSerThrGluArgLysThrProValLysLysLeuAlaProGlu 350
 |||||
 Db 301 AGGAAATCCCATTTTGAAAGTACAGAAAGAAAGACTCTGTTAAAAAGCTGCTCTGAA 360
 QY 351 PheSerLys 353
 |||||
 Db 361 TTTTCAAAA 369

RESULT 49
 A1457684 366 bp mRNA linear EST 13-APR-1999
 LOCUS t149607.x1 Soares NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:2144796 3' similar to contains MER22.C2 MER22 repetitive
 element ;, mRNA sequence.

ACCESSION A1457684
 VERSION A1457684.1 GI:4310553
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE NIH-MGC http://www.ncbi.nlm.nih.gov/ncigap.
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1203 Std Error: 0.00
 Seq primer: -40UP from G1bco
 High quality sequence stop: 347.
 Location/Qualifiers

FEATURES

source

1..366
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2144796"
 /lab_host="DH10B"
 /clone_1ib="Soares NSF_F8_9W_OT_PA_P_S1"
 /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and as circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and clones: Soares NBHSF pool 1:
 309384-310919, 323208-325895 Soares NBHP pool 1:
 145032-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NB2HF8-9W pool 1:

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2005, 15:43:39 ; Search time 45 Seconds
(without alignments)
1018.545 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 614
Sequence: 1 MTKMSAVFKAAAPAGNP.....RSRCRLIQKQTPHRTCRK 614

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 513545 seqs, 74649064 residues

Word size: 0

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database: Issued Patents_AA.*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.3	64	US-09-513-999C-6282	Sequence 6282, Ap
2	8	1.3	70	US-09-621-976-4979	Sequence 4979, Ap
3	8	1.3	83	US-08-834-306-63	Sequence 63, Appl
4	8	1.3	83	US-08-993-674A-63	Sequence 63, Appl
5	8	1.3	83	US-09-256-976-63	Sequence 63, Appl
6	8	1.3	109	US-09-849-602-17	Sequence 17, Appl
7	8	1.3	115	US-09-270-767-58601	Sequence 58601, A
8	8	1.3	123	US-09-513-999C-6025	Sequence 6025, Ap
9	8	1.3	142	US-09-187-999-19	Sequence 19, Appl
10	8	1.3	146	US-09-621-976-4920	Sequence 4920, Ap
11	8	1.3	148	US-09-248-796A-19533	Sequence 19533, A
12	8	1.3	169	US-09-489-039A-8371	Sequence 8371, Ap
13	8	1.3	184	US-09-252-991A-22743	Sequence 22743, A
14	8	1.3	223	US-09-252-991A-26883	Sequence 26883, A
15	8	1.3	232	US-09-555-352-10	Sequence 10, Appl
16	8	1.3	240	US-09-270-767-44271	Sequence 44271, A
17	8	1.3	242	US-09-248-796A-21033	Sequence 21033, A
18	8	1.3	330	US-09-301-666A-10	Sequence 10, Appl
19	8	1.3	330	US-09-301-217-10	Sequence 10, Appl
20	8	1.3	347	US-09-270-767-43260	Sequence 43260, A
21	8	1.3	432	US-09-170-984-10	Sequence 10, Appl
22	8	1.3	501	US-09-328-352-6940	Sequence 6940, Ap
23	8	1.3	512	US-09-252-991A-31457	Sequence 31457, A
24	8	1.3	619	US-09-578-441-3	Sequence 4, Appl
25	8	1.3	620	US-09-578-441-4	Sequence 4, Appl
26	8	1.3	628	US-09-252-991A-18780	Sequence 18780, A
27	8	1.3	639	US-09-949-016-7028	Sequence 7028, Ap

28	8	1.3	664	US-09-949-016-8177	Sequence 8177, Ap
29	8	1.3	668	US-09-949-016-10511	Sequence 10511, A
30	8	1.3	698	US-08-941-445A-11	Sequence 11, Appl
31	8	1.3	783	US-09-252-991A-18035	Sequence 18035, A
32	8	1.3	1088	US-09-233-857-13	Sequence 13, Appl
33	8	1.3	1099	US-09-442-100-2	Sequence 2, Appl
34	8	1.3	1099	US-09-939-106-2	Sequence 2, Appl
35	8	1.3	1099	US-09-442-102-2	Sequence 2, Appl
36	8	1.3	1844	US-08-851-567B-53	Sequence 53, Appl
37	8	1.3	2504	US-08-851-567B-12	Sequence 12, Appl
38	8	1.3	2504	US-09-817-514A-8	Sequence 8, Appl
39	8	1.3	2860	US-08-826-267-2	Sequence 2, Appl
40	8	1.1	14	US-08-557-309B-62	Sequence 62, Appl
41	8	1.1	14	US-08-557-309B-63	Sequence 63, Appl
42	8	1.1	14	US-08-929-414-14	Sequence 14, Appl
43	8	1.1	14	US-08-929-414-15	Sequence 15, Appl
44	8	1.1	14	US-08-834-306-61	Sequence 61, Appl
45	8	1.1	14	US-08-834-306-62	Sequence 62, Appl
46	8	1.1	14	US-08-993-674A-61	Sequence 61, Appl
47	8	1.1	14	US-08-993-674A-62	Sequence 62, Appl
48	8	1.1	14	US-09-256-976-61	Sequence 61, Appl
49	8	1.1	14	US-09-256-976-62	Sequence 62, Appl
50	8	1.1	14	US-09-674-973A-208	Sequence 208, Ap
51	8	1.1	15	US-08-602-999A-385	Sequence 385, Ap
52	8	1.1	15	US-09-500-124-385	Sequence 385, Ap
53	8	1.1	18	US-09-701-947A-90	Sequence 90, Appl
54	8	1.1	20	US-09-428-082B-1048	Sequence 1048, Ap
55	8	1.1	21	US-08-403-379A-53	Sequence 3, Appl
56	8	1.1	21	US-08-557-309B-52	Sequence 52, Appl
57	8	1.1	21	US-08-557-309B-57	Sequence 57, Appl
58	8	1.1	21	US-08-557-309B-58	Sequence 58, Appl
59	8	1.1	21	US-08-929-414-3	Sequence 3, Appl
60	8	1.1	21	US-08-929-414-10	Sequence 10, Appl
61	8	1.1	21	US-08-929-414-11	Sequence 11, Appl
62	8	1.1	21	US-08-484-905-14	Sequence 14, Appl
63	8	1.1	21	US-08-481-985B-14	Sequence 14, Appl
64	8	1.1	21	US-08-851-749-1	Sequence 1, Appl
65	8	1.1	21	US-08-834-306-55	Sequence 55, Appl
66	8	1.1	21	US-08-834-306-56	Sequence 56, Appl
67	8	1.1	21	US-08-834-306-58	Sequence 58, Appl
68	8	1.1	21	US-08-834-306-59	Sequence 59, Appl
69	8	1.1	21	US-08-370-476-14	Sequence 14, Appl
70	8	1.1	21	US-08-993-674A-55	Sequence 55, Appl
71	8	1.1	21	US-08-993-674A-56	Sequence 56, Appl
72	8	1.1	21	US-08-993-674A-58	Sequence 58, Appl
73	8	1.1	21	US-09-248-588-46	Sequence 46, Appl
74	8	1.1	21	US-09-256-976-55	Sequence 55, Appl
75	8	1.1	21	US-09-256-976-56	Sequence 56, Appl
76	8	1.1	22	US-08-406-824A-24	Sequence 24, Appl
77	8	1.1	26	US-07-789-913-8	Sequence 8, Appl
78	8	1.1	26	US-07-789-913-19	Sequence 19, Appl
79	8	1.1	26	US-08-049-794-8	Sequence 8, Appl
80	8	1.1	26	US-08-049-794-19	Sequence 19, Appl
81	8	1.1	26	US-08-496-847-8	Sequence 8, Appl
82	8	1.1	26	US-08-496-847-19	Sequence 19, Appl
83	8	1.1	26	US-08-742-774-8	Sequence 8, Appl
84	8	1.1	26	US-08-742-774-19	Sequence 19, Appl
85	8	1.1	26	US-08-675-354-8	Sequence 8, Appl
86	8	1.1	26	US-08-675-354-19	Sequence 19, Appl
87	8	1.1	26	US-08-965-918-8	Sequence 8, Appl
88	8	1.1	26	US-08-965-918-19	Sequence 19, Appl
89	8	1.1	26	US-09-138-439-8	Sequence 8, Appl
90	8	1.1	26	US-09-138-439-19	Sequence 19, Appl
91	8	1.1	26	US-08-613-400A-8	Sequence 8, Appl
92	8	1.1	26	US-08-613-400A-19	Sequence 19, Appl
93	8	1.1	26	US-09-238-017-8	Sequence 8, Appl
94	8	1.1	26	US-09-238-017-19	Sequence 19, Appl
95	8	1.1	26	US-09-392-979A-8	Sequence 8, Appl
96	8	1.1	26	US-09-392-979A-19	Sequence 19, Appl
97	8	1.1	26	US-09-062-440-12	Sequence 12, Appl
98	8	1.1	29	US-09-712-495-12	Sequence 12, Appl
99	8	1.1	29	US-09-270-767-59068	Sequence 59068, A
100	7	1.1	4	US-09-270-767-51268	Sequence 61268, A

ALIGNMENTS

RESULT 1

```

US-09-513-999C-6282
; Sequence 6282, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 6282
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 6
; OTHER INFORMATION: Xaa=Phe or Val
US-09-513-999C-6282

```

```

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 64;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 416 FSSRIEDL 423

Db 24 FSSRIEDL 31

RESULT 2

```

US-09-621-976-4979
; Sequence 4979, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 4979
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-4979

```

```

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 70;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 10 AAAAPAG 17

Db 10 AAAAPAG 17

RESULT 3

```

US-08-834-306-63
; Sequence 63, Application US/08834306

```

```

; Patent No. 6054135
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

```

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/834,306

FILING DATE: 15-APR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.422C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:

LENGTH: 83 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-08-834-306-63

```

Query Match
Best Local Similarity 100.0%; Score 8; DB 3; Length 83;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 9 KAAAPAG 16

Db 56 KAAAPAG 63

RESULT 4

```

US-08-993-674A-63
; Sequence 63, Application US/08993674A
; Patent No. 628372
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

```

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/993,674A
;; FILING DATE: 18-DEC-1997
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: MAKI, David J.
;; REGISTRATION NUMBER: 31,392
;; REFERENCE/DOCKET NUMBER: 210121.422C2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 63:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 83 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
US-08-993-674A-63

Query Match 1.3%; Score 8; DB 3; Length 83;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 KAAAPAG 16
Db 56 KAAAPAG 63

RESULT 5
US-09-256-976-63
;; Sequence 63, Application US/09256976
;; Patent No. 6419933
;; GENERAL INFORMATION:
;; APPLICANT: Reed, Steven G.
;; APPLICANT: Skeiky, Yasir A.W.
;; APPLICANT: Lodes, Michael J.
;; APPLICANT: Houghton, Raymond L.
;; APPLICANT: Smith, John M.
;; APPLICANT: McNeill, Patricia D.
;; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
;; TITLE OF INVENTION: OF T. CRUZI INFECTION
;; FILE REFERENCE: 210121.422C3
;; CURRENT APPLICATION NUMBER: US/09/256,976
;; CURRENT FILING DATE: 1999-02-24
;; NUMBER OF SEQ ID NOS: 95
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 63
;; LENGTH: 83
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Tetrapeptide
;; OTHER INFORMATION: containing four immunoreactive T. cruzi epitopes
;; OTHER INFORMATION: PEP-2, TCD, TCE and TcLo1.2 in a linear sequence
US-09-256-976-63

Query Match 1.3%; Score 8; DB 4; Length 83;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 KAAAPAG 16
Db 56 KAAAPAG 63

RESULT 6
US-09-849-602-17
;; Sequence 17, Application US/09849602
;; Patent No. 6794501
;; GENERAL INFORMATION:
;; APPLICANT: Scanlan, Matthew J.
;; APPLICANT: Old, Lloyd J.
;; APPLICANT: Stockert, Elisabeth

;; APPLICANT: Chen, Yao-Tseng
;; TITLE OF INVENTION: Colon Cancer Antigen Panel
;; FILE REFERENCE: 10461/7105 (CRV)
;; CURRENT APPLICATION NUMBER: US/09/849,602
;; CURRENT FILING DATE: 2001-05-04
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 17
;; LENGTH: 109
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: UNSURE
;; LOCATION: (84)..(84)
;; OTHER INFORMATION: x = any amino acid
;; FEATURE:
;; NAME/KEY: UNSURE
;; LOCATION: (100)..(100)
;; OTHER INFORMATION: x = any amino acid
US-09-849-602-17

Query Match 1.3%; Score 8; DB 4; Length 109;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 ASPATAS 167
Db 102 ASPATAS 109

RESULT 7
US-09-270-767-58601
;; Sequence 58601, Application US/09270767
;; Patent No. 6703491
;; GENERAL INFORMATION:
;; APPLICANT: Homburger et al.
;; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;; FILE REFERENCE: File Reference: 7326-094
;; CURRENT APPLICATION NUMBER: US/09/270,767
;; CURRENT FILING DATE: 1999-03-17
;; NUMBER OF SEQ ID NOS: 62517
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 58601
;; LENGTH: 115
;; TYPE: PRT
;; ORGANISM: Drosophila melanogaster
;; FEATURE:
;; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-58601

Query Match 1.3%; Score 8; DB 4; Length 115;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 338 TERKPYK 345
Db 6 TERKPYK 13

RESULT 8
US-09-513-999C-6025
;; Sequence 6025, Application US/09513999C
;; Patent No. 6783861
;; GENERAL INFORMATION:
;; APPLICANT: Dumas Milne Edwards, J.B.
;; APPLICANT: Duclert, A.
;; APPLICANT: Giordano, J.Y.
;; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
;; FILE REFERENCE: 59.US2.REG
;; CURRENT APPLICATION NUMBER: US/09/513,999C
;; CURRENT FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26
 ; NUMBER OF SEQ ID NOS: 36681
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 6025
 ; LENGTH: 123
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-513-999C-6025

Query Match 1.3%; Score 8; DB 4; Length 123;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 416 FSEIEDL 423
 Db 51 FSEIEDL 58

RESULT 9
 US-09-187-999-19
 ; Sequence 19; Application US/09187999A
 ; Patent No. 6482646
 ; GENERAL INFORMATION:
 ; APPLICANT: Gindullis, Frank
 ; APPLICANT: Meier, Iris
 ; TITLE OF INVENTION: Plant Proteins that Interact with Nuclear Matrix
 ; TITLE OF INVENTION: Proteins and Function as Transcriptional Activators
 ; FILE REFERENCE: CL-1321
 ; CURRENT APPLICATION NUMBER: US/09/187,999A
 ; CURRENT FILING DATE: 1998-11-06
 ; NUMBER OF SEQ ID NOS: 39
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 19
 ; LENGTH: 142
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 US-09-187-999-19

Query Match 1.3%; Score 8; DB 4; Length 142;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17
 Db 13 AAAAPAG 20

RESULT 10
 US-09-621-976-4920
 ; Sequence 4920; Application US/09621976
 ; Patent No. 6639063
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J. B.
 ; APPLICANT: Jobert, S.
 ; APPLICANT: Giordano, J. Y.
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 ; FILE REFERENCE: GENSET.054PR2
 ; CURRENT APPLICATION NUMBER: US/09/621,976
 ; CURRENT FILING DATE: 2000-07-21
 ; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 4920
 ; LENGTH: 146
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: 3
 ; OTHER INFORMATION: Xaa = Ala, Asp, Glu, Gly, Val
 US-09-621-976-4920

Query Match 1.3%; Score 8; DB 4; Length 146;
 Best Local Similarity 100.0%; Pred. No. 30;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 416 FSEIEDL 423
 Db 50 FSEIEDL 57

RESULT 11
 US-09-248-796A-19533
 ; Sequence 19533; Application US/09248796A
 ; Patent No. 6747137
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstock et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
 ; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; CURRENT FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 60/074,725
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: US 60/096,409
 ; PRIOR FILING DATE: 1998-08-13
 ; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 19533
 ; LENGTH: 148
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 US-09-248-796A-19533

Query Match 1.3%; Score 8; DB 4; Length 148;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 QPPSPPL 445
 Db 47 QPPSPPL 54

RESULT 12
 US-09-489-039A-8371
 ; Sequence 8371; Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 8371
 ; LENGTH: 169
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (10), (11), (12), (13), (14), (15), (16), (17), (18), (19), (20), (21), (22)
 ; LOCATION: (23), (24), (25), (26), (27), (28), (29), (30), (31), (32), (33), (34), (35)
 ; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
 US-09-489-039A-8371

Query Match 1.3%; Score 8; DB 4; Length 169;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 PPSPLPL 446
 Db 39 PPSPLPL 46

RESULT 13

US-09-252-991A-22743
; Sequence 22743, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22743
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22743

Query Match 1.3%; Score 8; DB 4; Length 184;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GGEPAANG 122
DB 36 GGEPAANG 43

RESULT 14
US-09-252-991A-26883
; Sequence 26883, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26883
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26883

Query Match 1.3%; Score 8; DB 4; Length 223;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 AASPATA 166
DB 31 AASPATA 38

RESULT 15
US-09-555-352-10
; Sequence 10, Application US/09555352
; Patent No. 6544779
; GENERAL INFORMATION:
; APPLICANT: Cichutek, Klaus
; TITLE OF INVENTION: PSEUDO-TYPE RETROVIRAL VECTORS WITH
; FILE REFERENCE: 11692-005001
; CURRENT APPLICATION NUMBER: US/09/555,352
; CURRENT FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: PCT/DE98/03542
; PRIOR FILING DATE: 1998-11-27
; PRIOR APPLICATION NUMBER: DE 197 52 855.4
; PRIOR FILING DATE: 1997-11-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Murine leukemia virus
US-09-555-352-10

Query Match 1.3%; Score 8; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 GGGSGSAS 216
DB 56 GGGSGSAS 63

RESULT 16
US-09-270-767-44271
; Sequence 44271, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44271
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44271

Query Match 1.3%; Score 8; DB 4; Length 240;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGSA 215
DB 65 GGGGSGSA 72

RESULT 17
US-09-248-796A-21033
; Sequence 21033, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 21033
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-21033

Query Match 1.3%; Score 8; DB 4; Length 242;

Best Local Similarity 100.0%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 EQOQOQLO 241
Db 97 EQOQOQLO 104

RESULT 18
US-09-301-666A-10

; Sequence 10, Application US/09301666A
; Patent No. 6495742
; GENERAL INFORMATION:
; APPLICANT: SHINOZAKI, Kazuko
; APPLICANT: KASUGA, Mie
; TITLE OF INVENTION: Genes Encoding Plant Transcription Factors
; FILE REFERENCE: 382.1028
; CURRENT APPLICATION NUMBER: US/09/301,666A
; CURRENT FILING DATE: 1999-04-28
; PRIOR APPLICATION NUMBER: JP228457/1998
; PRIOR FILING DATE: 1998-08-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-301-666A-10

Query Match 1.3%; Score 8; DB 4; Length 330;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 EQOQOQLO 241
Db 242 EQOQOQLO 249

RESULT 19
US-09-301-217-10

; Sequence 10, Application US/09301217
; Patent No. 6670528
; GENERAL INFORMATION:
; APPLICANT: Kazuko, SHINOZAKI
; APPLICANT: Mie, KASUGA
; TITLE OF INVENTION: Environmental Stress-tolerant Plant
; FILE REFERENCE: 382.1029
; CURRENT APPLICATION NUMBER: US/09/301,217
; CURRENT FILING DATE: 1999-04-28
; PRIOR APPLICATION NUMBER: JP292348/1998
; PRIOR FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-301-217-10

Query Match 1.3%; Score 8; DB 4; Length 330;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 EQOQOQLO 241
Db 242 EQOQOQLO 249

RESULT 20

US-09-270-767-43260
; Sequence 43260, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 43260
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-43260

Query Match 1.3%; Score 8; DB 4; Length 347;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 TERKTPVK 345
Db 238 TERKTPVK 245

RESULT 21

US-09-170-984-10
; Sequence 10, Application US/09170984
; Patent No. 6746867
; GENERAL INFORMATION:
; APPLICANT: Gillespie, Laura L.
; APPLICANT: Paterno, Gary D.
; TITLE OF INVENTION: Mammalian Mesoderm Induction Early Response
; FILE REFERENCE: 07468/017001
; CURRENT APPLICATION NUMBER: US/09/170,984
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-984-10

Query Match 1.3%; Score 8; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 416 FSSIEDL 423
Db 50 FSSIEDL 57

RESULT 22

US-09-328-352-6940
; Sequence 6940, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6940
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6940

Query Match 1.3%; Score 8; DB 4; Length 501;
Best Local Similarity 100.0%; Pred. No. 92;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 TGSIVAAA 146
|||||
Db 323 TGSIVAAA 330

RESULT 23
US-09-252-991A-31457
; Sequence 31457, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31457
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31457

Query Match 1.3%; Score 8; DB 4; Length 512;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17
|||||
Db 136 AAAAPAG 143

RESULT 24
US-09-578-441-3
; Sequence 3, Application US/09578441
; Patent No. 6562571
; GENERAL INFORMATION:
; APPLICANT: Wu, J.H. David
; APPLICANT: Omasa, Takeshi
; APPLICANT: Mantalaris, Athanasios
; APPLICANT: Chen, Yi-Guang
; TITLE OF INVENTION: HUMAN HEME-REGULATED INITIATION FACTOR 2. ALPHA KINASE
; FILE REFERENCE: 176/60571
; CURRENT APPLICATION NUMBER: US/09/578,441
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/135,713
; PRIOR FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-578-441-3

Query Match 1.3%; Score 8; DB 4; Length 619;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 EKEIEELK 251
|||||
Db 598 EKEIEELK 605

RESULT 25
US-09-578-441-4

; Sequence 4, Application US/09578441
; Patent No. 6562571
; GENERAL INFORMATION:
; APPLICANT: Wu, J.H. David
; APPLICANT: Omasa, Takeshi
; APPLICANT: Mantalaris, Athanasios
; APPLICANT: Chen, Yi-Guang
; APPLICANT: Teal, Ying-Chueh
; TITLE OF INVENTION: HUMAN HEME-REGULATED INITIATION FACTOR 2. ALPHA KINASE
; FILE REFERENCE: 176/60571
; CURRENT APPLICATION NUMBER: US/09/578,441
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/135,713
; PRIOR FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 620
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-578-441-4

Query Match 1.3%; Score 8; DB 4; Length 620;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 EKEIEELK 251
|||||
Db 599 EKEIEELK 606

RESULT 26
US-09-252-991A-18780
; Sequence 18780, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18780
; LENGTH: 628
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18780

Query Match 1.3%; Score 8; DB 4; Length 628;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 LPAGAPG 90
|||||
Db 287 LPAGAPG 294

RESULT 27
US-09-949-016-7028
; Sequence 7028, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14

```

; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7028
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7028

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 639;
Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 209 GGGGSGAS 216
DB 623 GGGGSGAS 630

RESULT 28
US-09-949-016-8177
; Sequence 8177, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8177
; LENGTH: 664
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8177

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 664;
Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 209 GGGGSGAS 216
DB 648 GGGGSGAS 655

RESULT 29
US-09-949-016-10511
; Sequence 10511, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR APPLICATION NUMBER: 60/231,498
```

```

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10511
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10511

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 668;
Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 165 TASDPAGP 172
DB 237 TASDPAGP 244

RESULT 30
US-08-941-445A-11
; Sequence 11, Application US/08941445A
; Patent No. 6107060
; GENERAL INFORMATION:
; APPLICANT: Keeling, Peter
; APPLICANT: Guan, Hanning
; TITLE OF INVENTION: Starch Encapsulation
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/941,445A
; FILING DATE: 30-SEP-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,855
; FILING DATE: 30-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Winner, Ellen P
; REGISTRATION NUMBER: 28,547
; REFERENCE/DOCKET NUMBER: 89-97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 698 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-941-445A-11

Query Match
Best Local Similarity 100.0%; Score 8; DB 3; Length 698;
Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 10 AAAAPAGG 17
DB 61 AAAAPAGG 68

RESULT 31
US-09-252-991A-18035
; Sequence 18035, Application US/09252991A
```



```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18035
; LENGTH: 783
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18035

Query Match          1.3%; Score 8; DB 4; Length 783;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      84 PAGAAGQ 91
      |||||
Db      547 PAGAAGQ 554

RESULT 32
US-09-233-857-13
; Sequence 13, Application US/09233857
; Patent No. 6495353
; GENERAL INFORMATION:
; APPLICANT: Plowman, Gregory
; APPLICANT: Flanagan, Peter
; TITLE OF INVENTION: HUMAN ORTHOLOGUES OF WART
; FILE REFERENCE: 239/251
; CURRENT APPLICATION NUMBER: US/09/233,857
; CURRENT FILING DATE: 1999-01-20
; EARLIER APPLICATION NUMBER: USSN 60/072,023
; EARLIER FILING DATE: 1998-01-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1088
; TYPE: PRT
; ORGANISM: HUMAN
US-09-233-857-13

Query Match          1.3%; Score 8; DB 4; Length 1088;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      235 QQQQQLQA 242
      |||||
Db      498 QQQQQLQA 505

RESULT 33
US-09-442-100-2
; Sequence 2, Application US/09442100
; Patent No. 6359193
; GENERAL INFORMATION:
; APPLICANT: Xu, Tian
; APPLICANT: Tao, Wufan
; APPLICANT: Wang, Weiyl
; APPLICANT: Zhang, Sheng
; APPLICANT: Yu, Wan
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
; TITLE OF INVENTION: GENES AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
```

```
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442,100
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/411,111
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6523-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1099 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-442-100-2

Query Match          1.3%; Score 8; DB 3; Length 1099;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      235 QQQQQLQA 242
      |||||
Db      498 QQQQQLQA 505

RESULT 34
US-08-939-106-2
; Sequence 2, Application US/08939106
; Patent No. 6559285
; GENERAL INFORMATION:
; APPLICANT: Yale University
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
; GENES AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/939,106
FILING DATE: 26-No. 6559285-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6523-007-228
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
```

TELEFAX: (212) 869-9741/8864
 TELEEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1099 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-08-939-106-2

Query Match 1.3%; Score 8; DB 4; Length 1099;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 QQQQQLQA 242
 Db 498 QQQQQLQA 505

RESULT 35
 US-09-442-102-2
 ; Sequence 2, Application US/09442102
 ; Patent No. 6630613
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Tian
 ; APPLICANT: Tao, Wufan
 ; APPLICANT: Wang, Weiyl
 ; APPLICANT: Zhang, Sheng
 ; APPLICANT: Yu, Man
 ; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
 ; TITLE OF INVENTION: GENES AND METHODS BASED THEREON
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/442,102
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/411,111
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mistrock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 6523-003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-9741/8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1099 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-09-442-102-2

Query Match 1.3%; Score 8; DB 4; Length 1099;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 235 QQQQQLQA 242

Db 498 QQQQQLQA 505

RESULT 36
 US-08-851-567B-53
 ; Sequence 53, Application US/08851567B
 ; Patent No. 6528484
 ; GENERAL INFORMATION:
 ; APPLICANT: Ensign, Jerald C
 ; APPLICANT: Bowen, David J
 ; APPLICANT: Petrell, James
 ; APPLICANT: Fatig, Raymond
 ; APPLICANT: Schoonover, Sue
 ; APPLICANT: French-Constant, Richard
 ; APPLICANT: Rocheleau, Thomas A.
 ; APPLICANT: Blackburn, Michael B.
 ; APPLICANT: Hey, Timothy D.
 ; APPLICANT: Merlo, Donald J.
 ; APPLICANT: Orr, Gregory L.
 ; APPLICANT: Roberts, Jean L.
 ; APPLICANT: Strickland, James A.
 ; APPLICANT: Guo, Lining
 ; APPLICANT: Cliche, Todd A.
 ; APPLICANT: Subhapinda, Kittisri
 ; TITLE OF INVENTION: Insecticidal Protein Toxins From Photobhabdus
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dow Agrosciences Patent Department
 ; STREET: 9330 Zionsville Road
 ; CITY: Indianapolis
 ; STATE: IN
 ; COUNTRY: US
 ; ZIP: 46268
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/851,567B
 ; FILING DATE: 05-MAY-1997
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/063,615
 ; FILING DATE: 18-MAY-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/395,497
 ; FILING DATE: 28-FEB-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/007,255
 ; FILING DATE: 06-NOV-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/608,423
 ; FILING DATE: 28-FEB-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/705,484
 ; FILING DATE: 28-AUG-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Seay, Nicholas J
 ; REGISTRATION NUMBER: 27386
 ; REFERENCE/DOCKET NUMBER: 960296.93804
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 608-251-5000
 ; TELEFAX: 608-251-9166
 ; INFORMATION FOR SEQ ID NO: 53:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1844 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-851-567B-53

Query Match 1.3%; Score 8; DB 4; Length 1844;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 ELSETSQT 314
DB 1588 ELSETSQT 1595

RESULT 37
US-08-851-567B-12

; Sequence 12, Application US/08851567B
; Patent No. 6528484

; GENERAL INFORMATION:

; APPLICANT: Ensign, Jerald C

; APPLICANT: Bowen, David J

; APPLICANT: Petell, James

; APPLICANT: Fatig, Raymond

; APPLICANT: Schooner, Sue

; APPLICANT: French-Constant, Richard

; APPLICANT: Rocheleau, Thomas A.

; APPLICANT: Blackburn, Michael B.

; APPLICANT: Hey, Timothy D.

; APPLICANT: Merlo, Donald J.

; APPLICANT: Orr, Gregory L.

; APPLICANT: Roberts, Jean L.

; APPLICANT: Strickland, James A.

; APPLICANT: Guo, Lining

; APPLICANT: Ciche, Todd A.

; APPLICANT: Sukhinda, Kitisri

; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus

; NUMBER OF SEQUENCES: 88

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dow Agrosciences Patent Department

; STREET: 9330 Zionsville Road

; CITY: Indianapolis

; STATE: IN

; COUNTRY: US

; ZIP: 46268

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/851,567B

; FILING DATE: 05-MAY-1997

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/063,615

; FILING DATE: 18-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/395,497

; FILING DATE: 28-FEB-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/007,255

; FILING DATE: 06-NOV-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/608,423

; FILING DATE: 28-FEB-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/705,484

; FILING DATE: 28-AUG-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Seay, Nicholas J

; REGISTRATION NUMBER: 27386

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 608-251-5000

; TELEFAX: 608-251-9166

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2504 amino acids

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
US-08-851-567B-12

Query Match 1.3%; Score 8; DB 4; Length 2504;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 ELSETSQT 314
DB 1675 ELSETSQT 1682

RESULT 38
US-09-817-514A-8

; Sequence 8, Application US/09817514A

; Patent No. 6639129

; GENERAL INFORMATION:

; APPLICANT: French-Constant, Richard

; APPLICANT: Bowen, David

; APPLICANT: Rocheleau, Thomas

; APPLICANT: Waterfield, Nicholas

; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS

; FILE REFERENCE: 61645

; CURRENT APPLICATION NUMBER: US/09/817,514A

; CURRENT FILING DATE: 2000-03-26

; PRIOR APPLICATION NUMBER: US 60/191806

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: Patentin version 3.0

; SEQ ID NO: 8

; LENGTH: 2504

; TYPE: PRT

; ORGANISM: Photorhabdus luminescens

; US-09-817-514A-8

Query Match 1.3%; Score 8; DB 4; Length 2504;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 ELSETSQT 314
DB 1675 ELSETSQT 1682

RESULT 39
US-08-826-267-2

; Sequence 2, Application US/08826267

; Patent No. 5994070

; GENERAL INFORMATION:

; APPLICANT: Streuli, Michel

; TITLE OF INVENTION: NO. 5994070e1 TRIO Molecules and Uses Related Thereto

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD

; STREET: 28 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/826,267

; FILING DATE: 1997

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/014,214

; FILING DATE: 27 MARCH (1996)

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragoras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2860 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-826-267-2

Query Match 1.3%; Score 8; DB 2; Length 2860;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
Db 2235 GGGGGSGA 2242

RESULT 40
US-08-557-309B-62
; Sequence 62, Application US/08557309B
; Patent No. 5916572
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,309B
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-557-309B-62

Query Match 1.1%; Score 7; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPA 15
Db 1 KAAAPA 7
```

```

;
; RESULT 41
US-08-557-309B-63
; Sequence 63, Application US/08557309B
; Patent No. 5916572
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,309B
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-557-309B-63

Query Match 1.1%; Score 7; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPA 15
Db 1 KAAAPA 7

RESULT 42
US-08-929-414-14
; Sequence 14, Application US/08929414
; Patent No. 5942403
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Haughton, Raymond
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/929,414
FILING DATE: 15-SEP-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.406C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-929-414-14

Query Match 1.1%; Score 7; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 9 KAAAPAA 15
Db 1 KAAAPAA 7

RESULT 43
US-08-929-414-15
Sequence 15, Application US/08929414
Patent No. 5942403
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Houghton, Raymond
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/929,414
FILING DATE: 15-SEP-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.406C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-929-414-15

Query Match 1.1%; Score 7; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 9 KAAAPAA 15
Db 1 KAAAPAA 7

RESULT 44
US-08-834-306-61
Sequence 61, Application US/08834306
Patent No. 6054135
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-61

Query Match 1.1%; Score 7; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 9 KAAAPAA 15
Db 1 KAAAPAA 7

RESULT 45
US-08-834-306-62
Sequence 62, Application US/08834306
Patent No. 6054135
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
US-08-834-306-62

STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-62

Query Match 1.1%; Score 7; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
|||||
Db 1 KAAAPAA 7

RESULT 46
US-08-993-674A-61
Sequence 61, Application US/08993674A
Patent No. 6228372
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-62

TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-61

Query Match 1.1%; Score 7; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
|||||
Db 1 KAAAPAA 7

RESULT 47
US-08-993-674A-62
Sequence 62, Application US/08993674A
Patent No. 6228372
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-62

Query Match 1.1%; Score 7; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
|||||
Db 1 KAAAPAA 7

RESULT 48
US-09-256-976-61

```
; Sequence 61, Application US/09256976
; Patent No. 6419933
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasar A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
; TITLE OF INVENTION: OF T. CRUZI INFECTION
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/09/256,976
; CURRENT FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TCE
; OTHER INFORMATION: antigenic epitope
US-09-256-976-61

Query Match          1.1%; Score 7; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
Db 1 KAAAPAA 7

RESULT 49
US-09-256-976-62
; Sequence 62, Application US/09256976
; Patent No. 6419933
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasar A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
; TITLE OF INVENTION: OF T. CRUZI INFECTION
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/09/256,976
; CURRENT FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TCE
; OTHER INFORMATION: antigenic epitope
US-09-256-976-62

Query Match          1.1%; Score 7; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
Db 1 KAAAPAA 7

RESULT 50
US-09-674-973A-208
; Sequence 208, Application US/09674973A
```

```
; Patent No. 6759046
; GENERAL INFORMATION:
; APPLICANT: No. 6759046sk Hydro ASA
; TITLE OF INVENTION: Peptides
; FILE REFERENCE: 26625-296
; CURRENT APPLICATION NUMBER: US/09/674,973A
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 459
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 208
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-674-973A-208

Query Match          1.1%; Score 7; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 PPSPLP 445
Db 8 PPSPLP 14

RESULT 51
US-08-602-999A-385
; Sequence 385, Application US/08602999A
; Patent No. 6184205
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: KAY, Brian K.
; APPLICANT: THORN, Judith M.
; APPLICANT: GUILIAM, Lawrence A.
; APPLICANT: DER, Channing J.
; APPLICANT: FOWLES, Dana M.
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF
; TITLE OF INVENTION: ISOLATING AND USING SAME
; NUMBER OF SEQUENCES: 467
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennile & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,999A
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 385:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-602-999A-385

Query Match          1.1%; Score 7; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 PPPLPLP 178
|||||
Db 7 PPPLPLP 13

RESULT 52
US-09-500-124-385
; Sequence 385, Application US/09500124
; Patent No. 6432920
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: KAY, Brian K.
; APPLICANT: THORN, Judith M.
; APPLICANT: OQUILLIAM, Lawrence A.
; APPLICANT: DER, Channing J.
; APPLICANT: FOWLIKES, Dana M.
; APPLICANT: RIDER, James E.
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF
; TITLE OF INVENTION: ISOLATING AND USING SAME
; NUMBER OF SEQUENCES: 467
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,124
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,999
; FILING DATE: 16-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 385:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-09-500-124-385

Query Match 1.1%; Score 7; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 PPPLPLP 178
|||||
Db 7 PPPLPLP 13

RESULT 53
US-09-701-947A-90
; Sequence 90, Application US/09701947A
; Patent No. 6818611
; GENERAL INFORMATION:
; APPLICANT: Altman, Elliot
; TITLE OF INVENTION: STABILIZED BIOACTIVE PEPTIDES AND METHODS OF

; TITLE OF INVENTION: IDENTIFICATION, SYNTHESIS AND USE
; FILE REFERENCE: 235.00010101
; CURRENT APPLICATION NUMBER: US/09/701,947A
; CURRENT FILING DATE: 2000-12-05
; PRIOR APPLICATION NUMBER: 60/104,013
; PRIOR FILING DATE: 1998-10-13
; PRIOR APPLICATION NUMBER: 60/112,150
; PRIOR FILING DATE: 1998-12-14
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 90
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: stabilized
; OTHER INFORMATION: peptide
US-09-701-947A-90

Query Match 1.1%; Score 7; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 KKDNEKE 278
|||||
Db 11 KKDNEKE 17

RESULT 54
US-09-428-082B-1048
; Sequence 1048, Application US/09428082B
; Patent No. 6660843
; GENERAL INFORMATION:
; APPLICANT: FEIGER, ULRICH
; APPLICANT: LIU, CHUAN-PA
; APPLICANT: CHEETHAM, JANET C.
; APPLICANT: BOONE, THOMAS CHARLES
; TITLE OF INVENTION: MODIFIED PEPTIDES AS THERAPEUTIC AGENTS
; FILE REFERENCE: A-527
; CURRENT APPLICATION NUMBER: US/09/428,082B
; CURRENT FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/105,371
; PRIOR FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 1133
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1048
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: IL-1 ANTAGONIST
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (20)-(20)
; OTHER INFORMATION: Fc domain attached at Position 20 of the C-terminus
US-09-428-082B-1048

Query Match 1.1%; Score 7; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 PLGGGGG 212
|||||
Db 14 PLGGGGG 20

RESULT 55
US-08-403-379A-3
; Sequence 3, Application US/08403379A
; Patent No. 5756662
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION


```

; TITLE OF INVENTION: OF T. CRUIZ INFECTION
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,379A
; FILING DATE: 14-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-403-379A-3
;
Query Match 1.1%; Score 7; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
Db 8 KAAAPAA 14

RESULT 56
US-08-557-309B-52
; Sequence 52, Application US/08557309B
; Patent No. 5916572
; GENERAL INFORMATION:
; APPLICANT: Reed,Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,309B
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-557-309B-52
;
Query Match 1.1%; Score 7; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
Db 8 KAAAPAA 14

RESULT 57
US-08-557-309B-57
; Sequence 57 Application US/08557309B
; Patent No. 5916572
; GENERAL INFORMATION:
; APPLICANT: Reed,Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,309B
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-557-309B-57
;
Query Match 1.1%; Score 7; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
Db 8 KAAAPAA 14

RESULT 58
US-08-557-309B-58

```

```
; Sequence 58, Application US/08557309B
; Patent No. 5916572
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,309B
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-557-309B-58
;
Query Match 1.1%; Score 7; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPPA 15
   |||||
Db 1 KAAAPPA 7

RESULT 59
US-08-929-414-3
; Sequence 3, Application US/08929414
; Patent No. 5942403
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Haughton, Raymond
; APPLICANT: Skelky, Yasir A.W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
; TITLE OF INVENTION: OF T. CRUZI INFECTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/929,414
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-929-414-3
```

```
Query Match 1.1%; Score 7; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 9 KAAAPPA 15
   |||||
Db 8 KAAAPPA 14
```

```
RESULT 60
US-08-929-414-10
; Sequence 10, Application US/08929414
; Patent No. 5942403
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Haughton, Raymond
; APPLICANT: Skelky, Yasir A.W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
; TITLE OF INVENTION: OF T. CRUZI INFECTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/929,414
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-929-414-10

Query Match 1.1%; Score 7; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	9	KAAAPPA	15
Db	8	KAAAPPA	14

RESULT 61
US-08-929-414-11
Sequence 11

```

Sequence 11, Application US/08929414
Patent No. 5942403
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Houghton, Raymond
APPLICANT: Skeiky, Yashir A.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
TITLE OF INVENTION: OF T. CRUZI INFECTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/929,414
FILING DATE: 15-SEP-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.406C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-929-414-11

```

Query Match	1.1%;	Score 7;	DB 2;	Length 21;
Best Local Similarity	100.0%;	Pred. No. 42;		
Matches	7;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
QY	9	KAAAPPA	15	
Db	1	KAAAPPA	7	

```

RESULT 62
US-08-484-905-14
; Sequence 14, Application US/08484905
; Patent No. 5976551
; GENERAL INFORMATION:
; APPLICANT: Motez, Estelle
; APPLICANT: Abastado, Jean-Pierre
; APPLICANT: Kourilsky, Philippe
; TITLE OF INVENTION: An Altered Major Histocompatibility
; TITLE OF INVENTION: Complex(MHC) Determinant and Methods for Using the
; TITLE OF INVENTION: Determinant
; NUMBER OF SEQUENCES: 127
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finegan, Henderson, Farbow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
;

```

```

STATE: D.C.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS-/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,905
FILING DATE: 07-JUNE-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/801,818
FILING DATE: 05-DEC-1991
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/792,473
FILING DATE: 15-NOV-1991
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 03495, 0106-03000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4400
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-484-905-14

```

Query	March	Similarity	1.1%;	Score	7;	DB	2;	length	21;	
	Best	Local	100.0%;	Pred.	No.	42;				
	Matches	7;	Conservative	0;	Mismatches			0;	Gaps	0;
Qy	208	GGGGGSG	214							
Db	6	GGGGGSG	12							

```

RESULT 63
US-08-851-749-1
; Sequence 1, Application US/08851749
; Patent No. 6001329
; GENERAL INFORMATION:
; APPLICANT: Buchsbaum, Donald J.
; APPLICANT: Vallera, Daniel A.
; APPLICANT: Blazer, Bruce R.
; TITLE OF INVENTION: Radiolabeled Fusion Toxins for Cancer Therapy
; FILE REFERENCE: D5654
; CURRENT APPLICATION NUMBER: US/08/851,749
; CURRENT FILING DATE: 1997-05-06
; EARLIER APPLICATION NUMBER: US 60/016,982
; EARLIER FILING DATE: 1996-05-06
; NUMBER OF SEQ ID NOS: 10
; SEQ ID NO 1
; LENGTH: 21
; TYPE: PRT
; ORGANISM: artificial sequence
FEATURE:
; OTHER INFORMATION: Amino acid sequence encoded by DNA linking mGM-CSF
; OTHER INFORMATION: to DT after restriction endonuclease digestion of
US-08-851-749-1

```

Query March	1.1%	Score 7	DB 3	Length 21
Best Local Similarity	100.0%	Pred. No. 42		
Matches	7	Conservative	0	Mismatches 0; Indels 0; Gaps 0;
OY	209	GGGGSGA	215	

Db 6 GGGGSGA 12

RESULT 64

US-08-481-985B-14
; Sequence 14, Application US/08481985B

; Patent No. 6011146

; GENERAL INFORMATION:

; APPLICANT: Mottez, Estelle

; APPLICANT: Abastado, Jean-Pierre

; APPLICANT: Kourileky, Philippe

; TITLE OF INVENTION: Altered Major Histocompatibility Complex

; NUMBER OF SEQUENCES: 148

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; STREET: 1300 I Street, N.W., Suite 700

; CITY: Washington

; STATE: D.C.

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/481,985B

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/801,818

; FILING DATE: 05-DEC-1991

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/792,473

; FILING DATE: 15-NOV-1991

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Meyers, Kenneth J.

; REGISTRATION NUMBER: 25,146

; REFERENCE/DOCKET NUMBER: 03495.0106-04000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-408-4000

; TELEFAX: 202-408-4400

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 21 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-481-985B-14

Query Match 1.1%; Score 7; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGG 214
Db 6 GGGGSGG 12

RESULT 65
US-08-834-306-55
; Sequence 55, Application US/08834306

; Patent No. 6054135

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T

NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/834,306

; FILING DATE: 15-APR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.422C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 55:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 21 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

US-08-834-306-55

Query Match 1.1%; Score 7; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPA 15
Db 8 KAAAPA 14

RESULT 66

US-08-834-306-56

; Sequence 56, Application US/08834306

; Patent No. 6054135

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T

; NUMBER OF SEQUENCES: 65

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/834,306

; FILING DATE: 15-APR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.422C1

; TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-56

Query Match 1.1%; Score 7; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
| | | | |
Db 1 KAAAPAA 7

RESULT 67
US-08-834-306-58
Sequence 58, Application US/08834306
Patent No. 6054135
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-58

Query Match 1.1%; Score 7; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
| | | | |
Db 1 KAAAPAA 7

RESULT 68
US-08-370-476-14
Sequence 14, Application US/08370476

Patent No. 6153408
GENERAL INFORMATION:
APPLICANT: Mottez, Estelle
APPLICANT: Abastado, Jean-Pierre
APPLICANT: Kourileky, Philippe
APPLICANT: Lome, Yu-Chun
APPLICANT: Ojcius, David
APPLICANT: Casrouge, Armand
TITLE OF INVENTION: Altered Major Histocompatibility Complex
NUMBER OF SEQUENCES: 127
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
City: Washington
STATE: D.C.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,476
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117,575
FILING DATE: 07-SEP-1993
APPLICATION NUMBER: US 08/072,787
FILING DATE: 06-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/801,818
FILING DATE: 05-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/792,473
FILING DATE: 15-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meyere, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 05243.0001-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-370-476-14

Query Match 1.1%; Score 7; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSG 214
| | | | |
Db 6 GGGGGSG 12

RESULT 69
US-08-993-674A-55
Sequence 55, Application US/08993674A
Patent No. 6228372
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-55

Query Match 1.1%; Score 7; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPPA 15
Db 8 KAAAPPA 14

RESULT 70
US-08-993-674A-56
Sequence 56, Application US/08993674A
Patent No. 6228372
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-56

Query Match 1.1%; Score 7; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPPA 15
Db 1 KAAAPPA 7

RESULT 71
US-08-993-674A-58
Sequence 58, Application US/08993674A
Patent No. 6228372
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-58

Query Match 1.1%; Score 7; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPPA 15
Db 1 KAAAPPA 7

```
RESULT 72
US-09-248-588-46
; Sequence 46, Application US/09248588
; Patent No. 6231864
; GENERAL INFORMATION:
; APPLICANT: Birkett, Ashley J.
; TITLE OF INVENTION: Strategically Modified Hepatitis B Core Proteins and
; TITLE OF INVENTION: their Derivatives
; FILE REFERENCE: SYN-101 4564/69529
; CURRENT APPLICATION NUMBER: US/09/248,588
; CURRENT FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: 60/074537
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: WO 97/18475
US-09-248-588-46

Query Match
Best Local Similarity 100.0%; DB 3; Length 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
Db 8 KAAAPAA 14

RESULT 73
US-09-256-976-55
; Sequence 55, Application US/09256976
; Patent No. 6419933
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
; TITLE OF INVENTION: OF T. CRUZI INFECTION
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/09/256,976
; CURRENT FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 55
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TCE antigenic
; OTHER INFORMATION: epitope
US-09-256-976-55

Query Match
Best Local Similarity 100.0%; DB 4; Length 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
Db 8 KAAAPAA 14

RESULT 74
US-09-256-976-56
; Sequence 56, Application US/09256976
```

```
; Patent No. 6419933
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
; TITLE OF INVENTION: OF T. CRUZI INFECTION
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/09/256,976
; CURRENT FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 56
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TCE antigenic
; OTHER INFORMATION: epitope
US-09-256-976-56

Query Match
Best Local Similarity 100.0%; DB 4; Length 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
Db 1 KAAAPAA 7

RESULT 75
US-09-256-976-58
; Sequence 58, Application US/09256976
; Patent No. 6419933
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
; TITLE OF INVENTION: OF T. CRUZI INFECTION
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/09/256,976
; CURRENT FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 58
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TCE
; OTHER INFORMATION: antigenic epitope
US-09-256-976-58

Query Match
Best Local Similarity 100.0%; DB 4; Length 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KAAAPAA 15
Db 1 KAAAPAA 7

Search completed: March 23, 2005, 15:53:34
Job time : 48 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2005, 15:52:12 ; Search time 141 Seconds
(without alignments)
1441.815 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 614
Sequence: 1 MTRSAVFKAAPAAAGAGNPE.....RSRCRLIOKQTPARTCK 614

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1407402 seqs, 33110923 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1407402

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Published Applications_AA.*
1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubppa/PCF_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppa/PCFUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubppa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubppa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubppa/US11_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	614	100.0	614	US-10-054-935-2	Sequence 2, Appl1
2	614	100.0	614	US-10-144-194A-66	Sequence 66, Appl
3	278	45.3	281	US-09-925-302-655	Sequence 655, App
4	278	45.3	281	US-09-925-302-655	Sequence 655, App
5	195	31.8	198	US-10-054-935-3	Sequence 4, Appl1
6	73	11.9	616	US-10-054-935-4	Sequence 53391, A
7	10	1.6	156	US-10-767-701-53591	Sequence 53391, A
8	10	1.6	1390	US-10-097-340-35	Sequence 103025, A
9	9	1.5	124	US-10-437-963-103025	Sequence 103025, A
10	9	1.5	231	US-10-437-963-134598	Sequence 134598, A
11	9	1.5	242	US-10-029-386-32982	Sequence 32982, A
12	9	1.5	297	US-10-425-114-64002	Sequence 64002, A
13	9	1.5	323	US-10-755-889-447	Sequence 447, App

14	9	1.5	594	US-10-437-963-107287	Sequence 107287, A
15	9	1.5	652	US-10-104-047-3364	Sequence 3364, App
16	8	1.3	10	US-09-572-270A-426	Sequence 426, App
17	8	1.3	10	US-09-572-270A-496	Sequence 496, App
18	8	1.3	10	US-09-572-270A-644	Sequence 644, App
19	8	1.3	10	US-09-572-270A-646	Sequence 646, App
20	8	1.3	10	US-09-572-270A-750	Sequence 750, App
21	8	1.3	15	US-10-056-583-94	Sequence 94, Appl
22	8	1.3	17	US-10-056-583-96	Sequence 96, Appl
23	8	1.3	79	US-10-424-599-273803	Sequence 273803, A
24	8	1.3	80	US-09-925-302-582	Sequence 582, App
25	8	1.3	80	US-09-925-302-582	Sequence 582, App
26	8	1.3	86	US-10-424-599-261952	Sequence 261952, A
27	8	1.3	90	US-10-437-963-174330	Sequence 174330, A
28	8	1.3	107	US-10-437-963-151998	Sequence 151998, A
29	8	1.3	109	US-09-849-602-17	Sequence 17, Appl
30	8	1.3	109	US-10-767-701-47105	Sequence 47105, A
31	8	1.3	109	US-10-767-701-47106	Sequence 47106, A
32	8	1.3	110	US-10-437-963-190130	Sequence 190130, A
33	8	1.3	112	US-10-158-057-232	Sequence 232, App
34	8	1.3	112	US-10-437-963-172495	Sequence 172495, A
35	8	1.3	115	US-10-425-114-48860	Sequence 48860, A
36	8	1.3	121	US-10-424-599-171746	Sequence 171746, A
37	8	1.3	127	US-10-424-599-215204	Sequence 215204, A
38	8	1.3	144	US-10-437-963-116911	Sequence 116911, A
39	8	1.3	149	US-10-437-963-137044	Sequence 137044, A
40	8	1.3	149	US-10-767-701-60716	Sequence 60716, A
41	8	1.3	162	US-10-425-114-46864	Sequence 46864, A
42	8	1.3	162	US-10-437-963-136138	Sequence 136138, A
43	8	1.3	162	US-10-437-963-169242	Sequence 169242, A
44	8	1.3	163	US-10-437-963-113136	Sequence 113136, A
45	8	1.3	165	US-10-425-114-72053	Sequence 72053, A
46	8	1.3	170	US-10-767-701-50943	Sequence 50943, A
47	8	1.3	173	US-10-425-114-37537	Sequence 37537, A
48	8	1.3	173	US-10-425-114-38104	Sequence 38104, A
49	8	1.3	174	US-10-437-963-105628	Sequence 105628, A
50	8	1.3	176	US-10-109-048-679	Sequence 679, App
51	8	1.3	183	US-10-425-114-61761	Sequence 61761, A
52	8	1.3	184	US-10-424-599-230617	Sequence 230617, A
53	8	1.3	185	US-10-767-701-61104	Sequence 61104, A
54	8	1.3	191	US-09-864-761-36885	Sequence 36885, A
55	8	1.3	196	US-10-437-963-148651	Sequence 148651, A
56	8	1.3	205	US-10-437-963-105048	Sequence 105048, A
57	8	1.3	208	US-10-029-386-33786	Sequence 33786, A
58	8	1.3	208	US-10-437-963-198586	Sequence 198586, A
59	8	1.3	216	US-10-424-599-215982	Sequence 215982, A
60	8	1.3	222	US-10-425-114-36874	Sequence 36874, A
61	8	1.3	242	US-10-425-114-67696	Sequence 67696, A
62	8	1.3	242	US-10-425-114-67778	Sequence 67778, A
63	8	1.3	251	US-10-767-701-33215	Sequence 33215, A
64	8	1.3	262	US-10-424-599-226303	Sequence 226303, A
65	8	1.3	263	US-10-437-963-164408	Sequence 164408, A
66	8	1.3	266	US-09-805-020-61	Sequence 61, Appl
67	8	1.3	269	US-10-437-963-147881	Sequence 147881, A
68	8	1.3	275	US-10-156-761-12256	Sequence 12256, A
69	8	1.3	292	US-09-805-020-60	Sequence 60, Appl
70	8	1.3	292	US-10-369-493-14460	Sequence 14460, A
71	8	1.3	292	US-10-369-493-14923	Sequence 14923, A
72	8	1.3	293	US-10-104-047-2088	Sequence 2088, App
73	8	1.3	304	US-09-927-091-2	Sequence 2, Appl1
74	8	1.3	314	US-10-424-599-185947	Sequence 185947, A
75	8	1.3	315	US-10-416-387-2	Sequence 2, Appl1
76	8	1.3	319	US-10-425-114-65361	Sequence 65361, A
77	8	1.3	319	US-10-425-114-68634	Sequence 68634, A
78	8	1.3	319	US-10-425-114-69464	Sequence 69464, A
79	8	1.3	330	US-10-266-487-10	Sequence 10, Appl
80	8	1.3	330	US-10-278-535-62	Sequence 62, Appl
81	8	1.3	330	US-10-412-699B-748	Sequence 748, App
82	8	1.3	336	US-10-437-963-127159	Sequence 127159, A
83	8	1.3	385	US-10-109-048-483	Sequence 483, App
84	8	1.3	378	US-10-424-599-254246	Sequence 254246, A
85	8	1.3	378	US-10-767-701-45744	Sequence 45744, A
86	8	1.3	381	US-10-767-701-36815	Sequence 36815, A

87 8 1.3 391 14 US-10-328-190-14 Sequence 14, Appl
88 8 1.3 397 15 US-10-425-114-62859 Sequence 62859, A
89 8 1.3 413 15 US-10-108-260A-4616 Sequence 4616, A
90 8 1.3 417 15 US-10-104-047-2563 Sequence 2563, A
91 8 1.3 419 15 US-10-425-114-47512 Sequence 47512, A
92 8 1.3 426 15 US-10-425-114-41019 Sequence 41019, A
93 8 1.3 445 14 US-10-156-761-9004 Sequence 9004, A
94 8 1.3 449 15 US-10-424-599-201859 Sequence 201859, A
95 8 1.3 454 15 US-10-425-114-63654 Sequence 63654, A
96 8 1.3 469 16 US-10-363-829-297 Sequence 297, A
97 8 1.3 470 15 US-10-283-122A-47712 Sequence 47712, A
98 8 1.3 472 15 US-10-369-493-20756 Sequence 20756, A
99 8 1.3 476 16 US-10-437-963-127603 Sequence 127603, A
100 8 1.3 487 16 US-10-437-963-137896 Sequence 137896, A

ALIGNMENTS

RESULT 1
; Sequence 2, Application US/10054935
; Publication No. US20030143546A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: BREAST CANCER TRANSCRIPTION FACTOR GENE AND USES
; FILE REFERENCE: 16U 107 R1
; CURRENT APPLICATION NUMBER: US/10/054, 935
; CURRENT FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 614
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-054-935-2

Query Match 100.0%; Score 614; DB 14; Length 614;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 614; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTRSAVFKAAPGAGNPEQRLDYERAAALGSPDEPGAARHFLPRHRKKEBPPLA 60
DB 1 MTRSAVFKAAPGAGNPEQRLDYERAAALGSPDEPGAARHFLPRHRKKEBPPLA 60
QY 61 SSQGSPPAPSPAGCGGKRGILLPAGAAFGQOEBSWGSVPLPCPPATKQAGIGEPAA 120
DB 61 SSQGSPPAPSPAGCGGKRGILLPAGAAFGQOEBSWGSVPLPCPPATKQAGIGEPAA 120
QY 121 AGAGCSPPRKYQAVLP1QIGSLVAAAKEPTPWAADKGAASPAATASDPAGPPPLP 180
DB 121 AGAGCSPPRKYQAVLP1QIGSLVAAAKEPTPWAADKGAASPAATASDPAGPPPLP 180
QY 181 PLAPATATGTLAASBGRKSMKSPPLGGGGSSGASQAACLKQILLQLDLIEQOQOOL 240
DB 181 PLAPATATGTLAASBGRKSMKSPPLGGGGSSGASQAACLKQILLQLDLIEQOQOOL 240
QY 241 QAKEKEIEELKSEBDTLARIERMERMOVLVKDNEKERHKLFGQYETERETELSEKI 300
DB 241 QAKEKEIEELKSEBDTLARIERMERMOVLVKDNEKERHKLFGQYETERETELSEKI 300
QY 301 KLECPBELSETSQTLPPKPFSCGRSGKGRKSPFGSTERKTPVKKLADPEFSKVTKTPK 360
DB 301 KLECPBELSETSQTLPPKPFSCGRSGKGRKSPFGSTERKTPVKKLADPEFSKVTKTPK 360
QY 361 HSPPIKEBPFGSLSEYCKRELASQETPEKPRSSVTPPRLSTPOKPSHPKKAASSSI 420
DB 361 HSPPIKEBPFGSLSEYCKRELASQETPEKPRSSVTPPRLSTPOKPSHPKKAASSSI 420
QY 421 EDLPYLSTTEMYLCRWHPPPSPPLPUBSSPKKEETVARCLMPSSVAGETSVLAVPSWD 480
DB 421 EDLPYLSTTEMYLCRWHPPPSPPLPUBSSPKKEETVARCLMPSSVAGETSVLAVPSWD 480

QY 481 HSEVPLRDPNPSDLLENLDDSVFSKRHAKLEDERRRKWD1QRIREOR11LQRLQRMKY 540
DB 481 HSEVPLRDPNPSDLLENLDDSVFSKRHAKLEDERRRKWD1QRIREOR11LQRLQRMKY 540
QY 541 KKGIOSEBEVTSFPPEPDVSLMITPFLPVVAFGRPLPKLTPONFELPWIDERSRCL 600
DB 541 KKGIOSEBEVTSFPPEPDVSLMITPFLPVVAFGRPLPKLTPONFELPWIDERSRCL 600
QY 601 EI0KKQTPHRTCRK 614
DB 601 EI0KKQTPHRTCRK 614

RESULT 2
US-10-144-194A-66
; Sequence 66, Application US/10144194A
; Publication No. US20030215809A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Breast Cancer Genes
; FILE REFERENCE: 3U 103 R1
; CURRENT APPLICATION NUMBER: US/10/144, 194A
; CURRENT FILING DATE: 2002-06-12
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 66
; LENGTH: 614
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-144-194A-66

Query Match 100.0%; Score 614; DB 15; Length 614;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 614; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTRSAVFKAAPGAGNPEQRLDYERAAALGSPDEPGAARHFLPRHRKKEBPPLA 60
DB 1 MTRSAVFKAAPGAGNPEQRLDYERAAALGSPDEPGAARHFLPRHRKKEBPPLA 60
QY 61 SSQGSPPAPSPAGCGGKRGILLPAGAAFGQOEBSWGSVPLPCPPATKQAGIGEPAA 120
DB 61 SSQGSPPAPSPAGCGGKRGILLPAGAAFGQOEBSWGSVPLPCPPATKQAGIGEPAA 120
QY 121 AGAGCSPPRKYQAVLP1QIGSLVAAAKEPTPWAADKGAASPAATASDPAGPPPLP 180
DB 121 AGAGCSPPRKYQAVLP1QIGSLVAAAKEPTPWAADKGAASPAATASDPAGPPPLP 180
QY 181 PLAPATATGTLAASBGRKSMKSPPLGGGGSSGASQAACLKQILLQLDLIEQOQOOL 240
DB 181 PLAPATATGTLAASBGRKSMKSPPLGGGGSSGASQAACLKQILLQLDLIEQOQOOL 240
QY 241 QAKEKEIEELKSEBDTLARIERMERMOVLVKDNEKERHKLFGQYETERETELSEKI 300
DB 241 QAKEKEIEELKSEBDTLARIERMERMOVLVKDNEKERHKLFGQYETERETELSEKI 300
QY 301 KLECPBELSETSQTLPPKPFSCGRSGKGRKSPFGSTERKTPVKKLADPEFSKVTKTPK 360
DB 301 KLECPBELSETSQTLPPKPFSCGRSGKGRKSPFGSTERKTPVKKLADPEFSKVTKTPK 360
QY 361 HSPPIKEBPFGSLSEYCKRELASQETPEKPRSSVTPPRLSTPOKPSHPKKAASSSI 420
DB 361 HSPPIKEBPFGSLSEYCKRELASQETPEKPRSSVTPPRLSTPOKPSHPKKAASSSI 420
QY 421 EDLPYLSTTEMYLCRWHPPPSPPLPUBSSPKKEETVARCLMPSSVAGETSVLAVPSWD 480
DB 421 EDLPYLSTTEMYLCRWHPPPSPPLPUBSSPKKEETVARCLMPSSVAGETSVLAVPSWD 480
QY 481 HSEVPLRDPNPSDLLENLDDSVFSKRHAKLEDERRRKWD1QRIREOR11LQRLQRMKY 540
DB 481 HSEVPLRDPNPSDLLENLDDSVFSKRHAKLEDERRRKWD1QRIREOR11LQRLQRMKY 540
QY 541 KKGIOSEBEVTSFPPEPDVSLMITPFLPVVAFGRPLPKLTPONFELPWIDERSRCL 600
DB 541 KKGIOSEBEVTSFPPEPDVSLMITPFLPVVAFGRPLPKLTPONFELPWIDERSRCL 600

Db 541 KKGIOSEBEPVTSFPEPDVDSLMTTPPLPVVAFGRLPKLTPONFELWDERGRCL 600
QY 601 EIQKQTPHRTCK 614
Db 601 EIQKQTPHRTCK 614

RESULT 3
US-09-925-302-655
; Sequence 655, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 655
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-655

Query Match 45.3%; Score 278; DB 9; Length 281;
Best Local Similarity 100.0%; Pred. No. 3,4e-232;
Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 181 PPLAPATAGTTLAASBGRWKSMPKSPGGGGGASQAACLKQILLDLIEQQOOL 240
Db 1 PPLAPATAGTTLAASBGRWKSMPKSPGGGGGASQAACLKQILLDLIEQQOOL 60
QY 241 QAKKEIEBELKSRDITLARIEMERRMOLVKDNEKERHKLFGYETERRETESEKI 300
Db 61 QAKKEIEBELKSRDITLARIEMERRMOLVKDNEKERHKLFGYETERRETESEKI 120
QY 301 KLECQELSETSQTLPPKPSGSGSGKHRSPPGSTERKTPVKCLAPFSVKTTPK 360
Db 121 KLECQELSETSQTLPPKPSGSGSGKHRSPPGSTERKTPVKCLAPFSVKTTPK 180
QY 361 HSPKEPCGSLSEYTCRKLRSQETPEKPRSSVDTPPRLSTPQKGPSTHPKKAFFSEI 420
Db 181 HSPKEPCGSLSEYTCRKLRSQETPEKPRSSVDTPPRLSTPQKGPSTHPKKAFFSEI 240
QY 421 EDLPYSTTEMYLCRWHPPPSPPLPRESSPKKEETVA 458
Db 241 EDLPYSTTEMYLCRWHPPPSPPLPRESSPKKEETVA 278

RESULT 4
US-09-925-302-655
; Sequence 655, Application US/09925302
; Publication No. US20030064072A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 655
; LENGTH: 281
; TYPE: PRT

; ORGANISM: Homo sapiens
US-09-925-302-655

Query Match 45.3%; Score 278; DB 10; Length 281;
Best Local Similarity 100.0%; Pred. No. 3,4e-232;
Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 181 PPLAPATAGTTLAASBGRWKSMPKSPGGGGGASQAACLKQILLDLIEQQOOL 240
Db 1 PPLAPATAGTTLAASBGRWKSMPKSPGGGGGASQAACLKQILLDLIEQQOOL 60
QY 241 QAKKEIEBELKSRDITLARIEMERRMOLVKDNEKERHKLFGYETERRETESEKI 300
Db 61 QAKKEIEBELKSRDITLARIEMERRMOLVKDNEKERHKLFGYETERRETESEKI 120
QY 301 KLECQELSETSQTLPPKPSGSGSGKHRSPPGSTERKTPVKCLAPFSVKTTPK 360
Db 121 KLECQELSETSQTLPPKPSGSGSGKHRSPPGSTERKTPVKCLAPFSVKTTPK 180
QY 361 HSPKEPCGSLSEYTCRKLRSQETPEKPRSSVDTPPRLSTPQKGPSTHPKKAFFSEI 420
Db 181 HSPKEPCGSLSEYTCRKLRSQETPEKPRSSVDTPPRLSTPQKGPSTHPKKAFFSEI 240
QY 421 EDLPYSTTEMYLCRWHPPPSPPLPRESSPKKEETVA 458
Db 241 EDLPYSTTEMYLCRWHPPPSPPLPRESSPKKEETVA 278

RESULT 5
US-10-054-935-3
; Sequence 3, Application US/10054935
; Publication No. US20030143546A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: BREAST CANCER TRANSCRIPTION FACTOR GENE AND USES
; FILE REFERENCE: 16U 107 R1
; CURRENT APPLICATION NUMBER: US/10/054,935
; CURRENT FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-054-935-3

Query Match 31.8%; Score 195; DB 14; Length 198;
Best Local Similarity 100.0%; Pred. No. 2e-160;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 MERRMOLVKDNEKERHKLFGYETERRETESEKIKLECQELSETSQTLPPKPSG 323
Db 1 MERRMOLVKDNEKERHKLFGYETERRETESEKIKLECQELSETSQTLPPKPSG 60
QY 324 RSGGKHRSPPGSTERKTPVKCLAPFSVKTTPKHSPIKEPCGSLSEYTCRKLRS 383
Db 61 RSGGKHRSPPGSTERKTPVKCLAPFSVKTTPKHSPIKEPCGSLSEYTCRKLRS 120
QY 384 QETPEKPRSSVDTPPRLSTPQKGPSTHPKKAFFSEIEDLPYSTTEMYLCRWHPPPSP 443
Db 121 QETPEKPRSSVDTPPRLSTPQKGPSTHPKKAFFSEIEDLPYSTTEMYLCRWHPPPSP 180
QY 444 LPLRESSPKKEETVA 458
Db 181 LPLRESSPKKEETVA 195

RESULT 6
US-10-054-935-4
; Sequence 4, Application US/10054935
; Publication No. US20030143546A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc

```

; TITLE OF INVENTION: BREAST CANCER TRANSCRIPTION FACTOR GENE AND USES
; FILE REFERENCE: 16U 107 R1
; CURRENT APPLICATION NUMBER: US/10/054, 935
; CURRENT FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 616
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-054-935-4

Query Match      11.9%; Score 73; DB 14; Length 616;
Best Local Similarity 100.0%; Pred. No. 2.6e-54;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 EDLPYLTTEMYLCWHQPPSPPLPLRSESSPKKEETVACLMPSVAGETSVLAIVPSWRD 480
DB 423 EDLPYLTTEMYLCWHQPPSPPLPLRSESSPKKEETVACLMPSVAGETSVLAIVPSWRD 482

QY 481 HSEVPLRDPNPSD 493
DB 483 HSEVPLRDPNPSD 495

RESULT 7
US-10-767-701-53391
; Sequence 53391, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5335)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 53391
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 13390462.pep
US-10-767-701-53391

Query Match      1.6%; Score 10; DB 16; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 PPLPLPGPP 181
DB 18 PPLPLPGPP 27

RESULT 8
US-10-097-340-35
; Sequence 35, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNANARAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
```

```

; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 1390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-35

Query Match      1.6%; Score 10; DB 14; Length 1390;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGASS 217
DB 99 GGGGSGGASS 108

RESULT 9
US-10-437-963-103025
; Sequence 103025, Application US/10437963
; Publication No. US20040123543A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Bardazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 103025
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_100495C.1.pep
US-10-437-963-103025

Query Match      1.5%; Score 9; DB 16; Length 124;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 PPPPLPLR 447
```

Db 107 PPPSPLPLR 115

RESULT 10

```

; Sequence 134598, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ. ID NOS: 204966
; SEQ. ID NO 134598
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_36356C.1.pep
US-10-437-963-134598

```

Query Match 1.5%; Score 9; DB 16; Length 211;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 PPPSPLPLR 447
 Db 222 PPPSPLPLR 230

RESULT 11

```

US-10-029-386-32982
; Sequence 32982, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ. ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ. ID NO 32982
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000287.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.88
; OTHER INFORMATION: SWISSPROT HIT: Q13516, EVALU8 2.00e-74
US-10-029-386-32982

```

Query Match 1.5%; Score 9; DB 14; Length 242;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGAS 216
 Db 193 GGGGGSGAS 201

RESULT 12

```

US-10-425-114-64002
; Sequence 64002, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 73128
; SEQ. ID NO 64002
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3245-269-H7_F11.pep
US-10-425-114-64002

```

Query Match 1.5%; Score 9; DB 15; Length 297;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LGGGGSGA 215
 Db 37 LGGGGSGA 45

RESULT 13

```

US-10-755-889-447
; Sequence 447, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ. ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ. ID NO 447
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-447

```

Query Match 1.5%; Score 9; DB 16; Length 323;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGAS 216
 Db 274 GGGGGSGAS 282

RESULT 14

```

US-10-437-963-107287
; Sequence 107287, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

```

```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barabazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 107287
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(594)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_11654C.1.pep
; US-10-437-963-107287

Query Match
Best Local Similarity 1.5%; Score 9; DB 16; Length 594;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 GGASPAATA 166
Db 565 GGASPAATA 573

RESULT 15
US-10-104-047-3364
; Sequence 3364, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3364
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-104-047-3364

Query Match
Best Local Similarity 1.5%; Score 9; DB 15; Length 652;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGSGAS 216
Db 562 GGGGSGSGAS 570

RESULT 16
US-09-572-270A-426
; Sequence 426, Application US/09572270A
; Publication No. US20030148368A1
; GENERAL INFORMATION:
; APPLICANT: Proteom Ltd
; TITLE OF INVENTION: Inter- complementary peptide listing
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/572,270A
; CURRENT FILING DATE: 2000-05-17
```

```

; NUMBER OF SEQ ID NOS: 1144
; SOFTWARE: ProPatent version 1.0
; SEQ ID NO 426
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
; OTHER INFORMATION: Sequence located in HAT9. at 232-241 and may interact with
; US-09-572-270A-426

Query Match
Best Local Similarity 1.3%; Score 8; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGA 215
Db 3 GGGGSGGA 10

RESULT 17
US-09-572-270A-496
; Sequence 496, Application US/09572270A
; Publication No. US20030148368A1
; GENERAL INFORMATION:
; APPLICANT: Proteom Ltd
; TITLE OF INVENTION: Inter- complementary peptide listing
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/572,270A
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 1144
; SOFTWARE: ProPatent version 1.0
; SEQ ID NO 496
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
; OTHER INFORMATION: Sequence located in HAT9. at 234-243 and may interact with
; US-09-572-270A-496

Query Match
Best Local Similarity 1.3%; Score 8; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGA 215
Db 1 GGGGSGGA 8

RESULT 18
US-09-572-270A-644
; Sequence 644, Application US/09572270A
; Publication No. US20030148368A1
; GENERAL INFORMATION:
; APPLICANT: Proteom Ltd
; TITLE OF INVENTION: Inter- complementary peptide listing
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/572,270A
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 1144
; SOFTWARE: ProPatent version 1.0
; SEQ ID NO 644
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
; OTHER INFORMATION: Sequence located in HAT9. at 234-243 and may interact with
; US-09-572-270A-644

Query Match
Best Local Similarity 1.3%; Score 8; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGA 215
Db 1 GGGGSGGA 8
```

```
RESULT 19
; US-09-572-270A-646
; Sequence 646, Application US/09572270A
; Publication No. US20030148368A1
; GENERAL INFORMATION:
; APPLICANT: Proteom Ltd
; TITLE OF INVENTION: Inter- complementary peptide listing
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/572,270A
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 1144
; SOFTWARE: Procpatent version 1.0
; SEQ ID NO 646
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
; OTHER INFORMATION: Sequence located in HAT9. at 233-242 and may interact with
US-09-572-270A-646

Query Match
1.3%; Score 8; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
DB 2 GGGGGSGA 9

RESULT 20
; US-09-572-270A-750
; Sequence 750, Application US/09572270A
; Publication No. US20030148368A1
; GENERAL INFORMATION:
; APPLICANT: Proteom Ltd
; TITLE OF INVENTION: Inter- complementary peptide listing
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/572,270A
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 1144
; SOFTWARE: Procpatent version 1.0
; SEQ ID NO 750
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
; OTHER INFORMATION: Sequence located in HAT9. at 233-242 and may interact with
US-09-572-270A-750

Query Match
1.3%; Score 8; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
DB 2 GGGGGSGA 9

RESULT 21
; US-10-056-583-94
; Sequence 94, Application US/10056583
; Publication No. US20030064915A1
; GENERAL INFORMATION:
; APPLICANT: Presidents and Fellows of Harvard College
; APPLICANT: Strominger, Jack L.
; APPLICANT: Fridkis-Hareli, Masha
; TITLE OF INVENTION: THERAPEUTIC PEPTIDES FOR DEMYELINATING
; FILE REFERENCE: 24655-017
; CURRENT APPLICATION NUMBER: US/10/056,583
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/263,569
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 99
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The peptide was designed and synthesized.
US-10-056-583-94

Query Match
1.3%; Score 8; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FKAAAPAA 15
DB 8 FKAAAPAA 15

RESULT 22
; US-10-056-583-96
; Sequence 96, Application US/10056583
; Publication No. US20030064915A1
; GENERAL INFORMATION:
; APPLICANT: Presidents and Fellows of Harvard College
; APPLICANT: Strominger, Jack L.
; APPLICANT: Fridkis-Hareli, Masha
; TITLE OF INVENTION: THERAPEUTIC PEPTIDES FOR DEMYELINATING
; FILE REFERENCE: 24655-017
; CURRENT APPLICATION NUMBER: US/10/056,583
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/263,569
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The peptide was designed and synthesized.
US-10-056-583-96

Query Match
1.3%; Score 8; DB 14; Length 17;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FKAAAPAA 15
DB 10 FKAAAPAA 17

RESULT 23
; US-10-424-599-273803
; Sequence 273803, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 273803
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
```

```
LOCATION: (1)..(79)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_89265C.1.pep
US-10-424-599-273803
```

```
Query Match 1.3%; Score 8; DB 15; Length 79;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 386 TPEKPRSS 393
DB 23 TPEKPRSS 30
```

RESULT 24

```
US-09-925-302-582
Sequence 582, Application US/09925302
Patent No. US20020044941A1
```

```
GENERAL INFORMATION:
```

```
APPLICANT: Rosen et al.
```

```
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
```

```
FILE REFERENCE: PA104
```

```
CURRENT APPLICATION NUMBER: US/09/925,302
```

```
PRIOR FILING DATE: 2001-08-10
```

```
PRIOR APPLICATION NUMBER: PCT/US00/05918
```

```
PRIOR FILING DATE: 2000-03-08
```

```
PRIOR APPLICATION NUMBER: 60/124,270
```

```
PRIOR FILING DATE: 1999-03-12
```

```
NUMBER OF SEQ ID NOS: 896
```

```
SOFTWARE: Patentin Ver. 2.0
```

```
SEQ ID NO 582
```

```
LENGTH: 80
```

```
TYPE: PRT
```

```
ORGANISM: Homo sapiens
```

```
US-09-925-302-582
```

```
Query Match 1.3%; Score 8; DB 9; Length 80;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 101 PLPCPPPA 108
DB 50 PLPCPPPA 57
```

RESULT 25

```
US-09-925-302-582
Sequence 582, Application US/09925302
Publication No. US20030064072A9
```

```
GENERAL INFORMATION:
```

```
APPLICANT: Rosen et al.
```

```
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
```

```
FILE REFERENCE: PA104
```

```
CURRENT APPLICATION NUMBER: US/09/925,302
```

```
PRIOR FILING DATE: 2001-08-10
```

```
PRIOR APPLICATION NUMBER: PCT/US00/05918
```

```
PRIOR FILING DATE: 2000-03-08
```

```
PRIOR APPLICATION NUMBER: 60/124,270
```

```
PRIOR FILING DATE: 1999-03-12
```

```
NUMBER OF SEQ ID NOS: 896
```

```
SOFTWARE: Patentin Ver. 2.0
```

```
SEQ ID NO 582
```

```
LENGTH: 80
```

```
TYPE: PRT
```

```
ORGANISM: Homo sapiens
```

```
US-09-925-302-582
```

```
Query Match 1.3%; Score 8; DB 10; Length 80;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 101 PLPCPPPA 108
```

```
DB 50 PLPCPPPA 57
```

RESULT 26

```
US-10-424-599-261952
Sequence 261952, Application US/10424599
Publication No. US20040031072A1
```

```
GENERAL INFORMATION:
```

```
APPLICANT: La Rosa Thomas J
```

```
APPLICANT: Kovalic David K
```

```
APPLICANT: Zhou Yihua
```

```
APPLICANT: Cao Yongwei
```

```
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
```

```
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

```
FILE REFERENCE: 38-21(53223)B
```

```
CURRENT APPLICATION NUMBER: US/10/424,599
```

```
PRIOR FILING DATE: 2003-04-28
```

```
NUMBER OF SEQ ID NOS: 285684
```

```
SEQ ID NO 261952
```

```
LENGTH: 86
```

```
TYPE: PRT
```

```
ORGANISM: Glycine max
```

```
FEATURE:
```

```
OTHER INFORMATION: Clone ID: PAT_MRT3847_78565C.1.pep
```

```
US-10-424-599-261952
```

```
Query Match 1.3%; Score 8; DB 15; Length 86;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 234 EQQQQQQLQ 241
DB 37 EQQQQQQLQ 44
```

RESULT 27

```
US-10-437-963-174330
Sequence 174330, Application US/10437963
Publication No. US20040123343A1
```

```
GENERAL INFORMATION:
```

```
APPLICANT: La Rosa, Thomas J.
```

```
APPLICANT: Kovalic, David K.
```

```
APPLICANT: Zhou, Yihua
```

```
APPLICANT: Cao, Yongwei
```

```
APPLICANT: Wu, Wei
```

```
APPLICANT: Boukharov, Andrey A.
```

```
APPLICANT: Barbazuk, Brad
```

```
APPLICANT: Li, Ping
```

```
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
```

```
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

```
FILE REFERENCE: 38-21(53221)B
```

```
CURRENT APPLICATION NUMBER: US/10/437,963
```

```
PRIOR FILING DATE: 2003-05-14
```

```
NUMBER OF SEQ ID NOS: 204966
```

```
SEQ ID NO 174330
```

```
LENGTH: 90
```

```
TYPE: PRT
```

```
ORGANISM: Oryza sativa
```

```
FEATURE:
```

```
OTHER INFORMATION: Clone ID: PAT_MRT4530_72281C.1.pep
```

```
US-10-437-963-174330
```

```
Query Match 1.3%; Score 8; DB 16; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 205 SPLGGGGG 212
DB 49 SPLGGGGG 56
```

RESULT 28

US-10-437-963-15198
; Sequence 15198, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 15198
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_52092C.1.pep
US-10-437-963-15198

Query Match 1.3%; Score 8; DB 16; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 439 PPPSPPLP 446
DB 21 PPPSPPLP 28

RESULT 29
US-09-849-602-17
; Sequence 17, Application US/09849602
; Publication No. US20030165834A1
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Old, Lloyd J.
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Chen, Yao-Tseng
; TITLE OF INVENTION: Colon Cancer Antigen Panel
; FILE REFERENCE: 10461/7105(JRV)
; CURRENT APPLICATION NUMBER: US/09/849,602
; CURRENT FILING DATE: 2001-05-04
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (84)-(84)
; OTHER INFORMATION: x = any amino acid
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (100)-(100)
; OTHER INFORMATION: x = any amino acid
US-09-849-602-17

Query Match 1.3%; Score 8; DB 10; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 ASPATAS 167
DB 102 ASPATAS 109

RESULT 30
US-10-767-701-47105
; Sequence 47105, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 47105
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C198_2.pep
US-10-767-701-47105

Query Match 1.3%; Score 8; DB 16; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAAPAGG 17
DB 72 AAAAPAGG 79

RESULT 31
US-10-767-701-47106
; Sequence 47106, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 47106
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C198_1.pep
US-10-767-701-47106

Query Match 1.3%; Score 8; DB 16; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAAPAGG 17
DB 72 AAAAPAGG 79

RESULT 32
US-10-437-963-190130
; Sequence 190130, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.

```
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 190130
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(110)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_86573C.1.pep
US-10-437-963-190130
```

```
Query Match 1.3%; Score 8; DB 16; Length 110;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 208 GGGGSGGA 215
DB 99 GGGGSGGA 106
```

```
RESULT 33
US-10-158-057-232
; Sequence 232, Application US/10158057
; Publication No. US20040014039A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1205C1
; CURRENT APPLICATION NUMBER: US/10/158,057
; CURRENT FILING DATE: 2002-06-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 232
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-158-057-232
```

```
Query Match 1.3%; Score 8; DB 15; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 321 SCGRSGKG 328
DB 41 SCGRSGKG 48
```

```
RESULT 34
US-10-437-963-172495
; Sequence 172495, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
```

```
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 172495
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_70625C.1.pep
US-10-437-963-172495
```

```
Query Match 1.3%; Score 8; DB 16; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 208 GGGGSGGA 215
DB 48 GGGGSGGA 55
```

```
RESULT 35
US-10-425-114-48860
; Sequence 48860, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48860
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3600-034-Fl1_Fli.pep
US-10-425-114-48860
```

```
Query Match 1.3%; Score 8; DB 15; Length 115;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 158 GAASPAAT 165
DB 82 GAASPAAT 89
```

```
RESULT 36
US-10-424-599-171746
; Sequence 171746, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 171746
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Glycine max
```

```

; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(121)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT847_126100C.1.pep
US-10-424-599-171746

Query Match
Best Local Similarity 1.3%; Score 8; DB 15; Length 121;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 GPPPLPAP 185
DB 42 GPPPLPAP 49

RESULT 37
US-10-424-599-215204
; Sequence 215204, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 215204
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_36354C.1.pep
US-10-424-599-215204

Query Match
Best Local Similarity 1.3%; Score 8; DB 15; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 QLOAKEKE 246
DB 52 QLOAKEKE 59

RESULT 38
US-10-437-963-116911
; Sequence 116911, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 116911
; LENGTH: 144
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
```

```

; NAME/KEY: unsure
; LOCATION: (1)..(144)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_20368C.1.pep
US-10-437-963-116911

Query Match
Best Local Similarity 1.3%; Score 8; DB 16; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 GPPPLPAP 178
DB 86 GPPPLPAP 93

RESULT 39
US-10-437-963-137044
; Sequence 137044, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 137044
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_38565C.1.pep
US-10-437-963-137044

Query Match
Best Local Similarity 1.3%; Score 8; DB 16; Length 149;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LGGGGGSG 214
DB 140 LGGGGGSG 147

RESULT 40
US-10-767-701-60716
; Sequence 60716, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 60716
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 9296298.pep
US-10-767-701-60716
```

Query Match 1.3%; Score 8; DB 16; Length 149;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 GGGLLLP 85
Db 10 GGGLLLP 17

RESULT 41

US-10-425-114-46864
; Sequence 46864, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46864
; LENGTH: 162
; TYPE: PR1
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701163276_FLI.pep
US-10-425-114-46864

Query Match 1.3%; Score 8; DB 15; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17
Db 125 AAAAPAG 132

RESULT 42

US-10-437-963-136138
; Sequence 136138, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 136138
; LENGTH: 162
; TYPE: PR1
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37746C.1.pep
US-10-437-963-136138

Query Match 1.3%; Score 8; DB 16; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17
Db 64 AAAAPAG 71

RESULT 43

US-10-437-963-169242
; Sequence 169242, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 169242
; LENGTH: 162
; TYPE: PR1
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(162)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_67680C.1.pep
US-10-437-963-169242

Query Match 1.3%; Score 8; DB 16; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 QPPSPLP 445
Db 65 QPPSPLP 72

RESULT 44

US-10-437-963-113136
; Sequence 113136, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 113136
; LENGTH: 163
; TYPE: PR1
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_16953C.1.pep
US-10-437-963-113136

Query Match 1.3%; Score 8; DB 16; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
DB 96 AATASDPA 103

RESULT 45
US-10-425-114-72053

; Sequence 72053, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72053
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB36-008-A5_FLI.pep
US-10-425-114-72053

Query Match 1.3%; Score 8; DB 15; Length 165;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGA 215
DB 2 GGGGSGA 9

RESULT 46
US-10-767-701-50943

; Sequence 50943, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 50943
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3480-045-PL-K1-A6.pep
US-10-767-701-50943

Query Match 1.3%; Score 8; DB 16; Length 170;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LGGGGSG 214
DB 116 LGGGGSG 123

RESULT 47
US-10-425-114-37537

; Sequence 37537, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 37537
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB22-027-G7_FLI.pep
US-10-425-114-37537

Query Match 1.3%; Score 8; DB 15; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGA 215
DB 133 GGGGSGA 140

RESULT 48
US-10-425-114-38104

; Sequence 38104, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 38104
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3069-003-H10_FLI.pep
US-10-425-114-38104

Query Match 1.3%; Score 8; DB 15; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 LPAGAAG 90
DB 1 LPAGAAG 8

RESULT 49
US-10-437-963-105628
; Sequence 105628, Application US/10437963

```
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 105628
LENGTH: 174
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_102854C.1.pep
US-10-437-963-105628
```

```
Query Match
1.3%; Score 8; DB 16; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 438 QPPSPPLP 445
DB 85 QPPSPPLP 92
```

```
RESULT 50
US-10-109-048-679
Sequence 679, Application US/10109048
Publication No. US20040107461A1
GENERAL INFORMATION:
APPLICANT: COMMURI, PADMA
APPLICANT: KEELING, PETER L.
APPLICANT: RAMIREZ, NONA
APPLICANT: MCKEAN, ANGELA
APPLICANT: GAO, ZHONG
APPLICANT: GUAN, HANPING
TITLE OF INVENTION: GLUCAN CHAIN LENGTH DOMAINS
FILE REFERENCE: 2461-76
CURRENT APPLICATION NUMBER: US/10/109,048
CURRENT FILING DATE: 2003-03-04
PRIOR APPLICATION NUMBER: 60/279,720
PRIOR FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 1154
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 679
LENGTH: 176
TYPE: PRT
ORGANISM: Unknown Organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Accession No. 7489711
US-10-109-048-679
```

```
Query Match
1.3%; Score 8; DB 16; Length 176;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 10 AAAAPAGG 17
DB 61 AAAAPAGG 68
```

```
RESULT 51
US-10-425-114-61761
Sequence 61761, Application US/10425114
Publication No. US20040034888A1
```

```
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaka, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 61761
LENGTH: 183
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3150-058-G3_F11.pep
US-10-425-114-61761
```

```
Query Match
1.3%; Score 8; DB 15; Length 183;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 208 GGGGSGGA 215
DB 20 GGGGSGGA 27
```

```
RESULT 52
US-10-424-599-230617
Sequence 230617, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovalic, David K
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 230617
LENGTH: 184
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_50269C.1.pep
US-10-424-599-230617
```

```
Query Match
1.3%; Score 8; DB 15; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 581 KLTPONFE 588
DB 67 KLTPONFE 74
```

```
RESULT 53
US-10-767-701-61104
Sequence 61104, Application US/10767701
Publication No. US20040172684A1
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53535)B
```

;; CURRENT APPLICATION NUMBER: US/10/767,701
;; CURRENT FILING DATE: 2004-01-29
;; NUMBER OF SEQ ID NOS: 63128
;; SEQ ID NO 61104
;; LENGTH: 185
;; TYPE: PRT
;; ORGANISM: Sorghum bicolor
;; FEATURE:
;; OTHER INFORMATION: Clone ID: 9301971.pcp
US-10-767-701-61104

Query Match 1.3%; Score 8; DB 16; Length 185;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 GGAASPA 164
DB 92 GGAASPA 99

RESULT 54
US-09-864-761-36985
; Sequence 36985, Application US/09864761
; Patent No. US2002048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36985

;; LENGTH: 191
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL121756.6
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.96
;; OTHER INFORMATION: SWISSPROT HIT: Q05819, EVALU 7.80e+00
US-09-864-761-36985

Query Match 1.3%; Score 8; DB 9; Length 191;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LGGGGSG 214
DB 82 LGGGGSG 89

RESULT 55
US-10-437-963-148651
; Sequence 148651, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 148651
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_49060C.1.pcp
US-10-437-963-148651

Query Match 1.3%; Score 8; DB 16; Length 196;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 PGPPPLAP 185
DB 102 PGPPPLAP 109

RESULT 56
US-10-437-963-105048
; Sequence 105048, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B

```

; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 105048
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_102325C.1.pep
US-10-437-963-105048

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 205;
Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 EQQQQQQLQ 241
DB 117 EQQQQQQLQ 124

RESULT 57
US-10-029-386-33786
; Sequence 33786, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33786
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121756.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.5
US-10-029-386-33786

Query Match
Best Local Similarity 100.0%; Score 8; DB 14; Length 208;
Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LGGGGGSG 214
DB 82 LGGGGGSG 89

RESULT 58
US-10-437-963-198586
; Sequence 198586, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 198586
```

```

; LENGTH: 208
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_94230C.1.pep
US-10-437-963-198586

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 208;
Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 SPLGGGGG 212
DB 84 SPLGGGGG 91

RESULT 59
US-10-424-599-215982
; Sequence 215982, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 28564
; SEQ ID NO 215982
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_37061C.1.pep
US-10-424-599-215982

Query Match
Best Local Similarity 100.0%; Score 8; DB 15; Length 216;
Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 VLPPIQGS 141
DB 112 VLPPIQGS 119

RESULT 60
US-10-425-114-36974
; Sequence 36974, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 36974
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3170-045-C12_FLI.pep
US-10-425-114-36974
```


Query Match 1.3%; Score 8; DB 15; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 IEEKXSR 254
DB 66 IEEKXSR 73

RESULT 61
US-10-425-114-67696

; Sequence 67696, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 67696
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3592-069-E8_FLI.pep
US-10-425-114-67696

Query Match 1.3%; Score 8; DB 15; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 PEDPGAA 41
DB 61 PEDPGAA 68

RESULT 62

US-10-425-114-67778
; Sequence 67778, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 67778
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3632-025-H11_FLI.pep
US-10-425-114-67778

Query Match 1.3%; Score 8; DB 15; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 PEDPGAA 41

DB 61 PEDPGAA 68

RESULT 63
US-10-767-701-33215

; Sequence 33215, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 33215
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C27405_1.pep
US-10-767-701-33215

Query Match 1.3%; Score 8; DB 16; Length 251;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 EPAAGAG 124
DB 142 EPAAGAG 149

RESULT 64

US-10-424-599-226303
; Sequence 226303, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 226303
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_46381C.1.pep
US-10-424-599-226303

Query Match 1.3%; Score 8; DB 15; Length 262;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGA 215
DB 101 GGGGSGA 108

RESULT 65

US-10-437-963-164408
; Sequence 164408, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

```

; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 164408
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_6330C.1.pep
US-10-437-963-164408

```

```

Query Match
Best Local Similarity 1.3%; Score 8; DB 16; Length 263;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 AAAPAG 17
Db 167 AAAPAG 174

```

```

RESULT 66
US-09-805-020-61
; Sequence 61, Application US/09805020
; Publication No. US20020086384A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, Zurit
; TITLE OF INVENTION: SPLICE VARIANTS OF ONCOGENES
; FILE REFERENCE: 2786-0168P
; CURRENT APPLICATION NUMBER: US/09/805,020
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 61
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-020-61

```

```

Query Match
Best Local Similarity 1.3%; Score 8; DB 9; Length 266;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 165 TASPAGP 172
Db 202 TASPAGP 209

```

```

RESULT 67
US-10-437-963-147881
; Sequence 147881, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B

```

```

; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147881
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_48367C.1.pep
US-10-437-963-147881

```

```

Query Match
Best Local Similarity 1.3%; Score 8; DB 16; Length 269;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 234 EQQOQLO 241
Db 255 EQQOQLO 262

```

```

RESULT 68
US-10-156-761-12256
; Sequence 12256, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIRAMA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12256
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12256

```

```

Query Match
Best Local Similarity 1.3%; Score 8; DB 14; Length 275;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 159 AASPATA 166
Db 261 AASPATA 268

```

```

RESULT 69
US-09-805-020-60
; Sequence 60, Application US/09805020
; Publication No. US20020086384A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, Zurit
; TITLE OF INVENTION: SPLICE VARIANTS OF ONCOGENES
; FILE REFERENCE: 2786-0168P
; CURRENT APPLICATION NUMBER: US/09/805,020
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 60
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-020-60

```

Query Match 1.3%; Score 8; DB 9; Length 292;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 TASDPAG 172
|||||
DB 202 TASDPAG 209

RESULT 70
US-10-369-493-14460

; Sequence 14460, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14460
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14460

Query Match 1.3%; Score 8; DB 15; Length 292;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
|||||
DB 233 AATASDPA 240

RESULT 71
US-10-369-493-14923

; Sequence 14923, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14923
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14923

Query Match 1.3%; Score 8; DB 15; Length 292;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 AATASDPA 170
|||||

DB 233 AATASDPA 240

RESULT 72

US-10-104-047-2088
; Sequence 2088, Application US/10104047
; Publication No. US200302336392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US200302336392A1e1 full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2088
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2088

Query Match 1.3%; Score 8; DB 15; Length 293;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAG 17
|||||
DB 10 AAAAPAG 17

RESULT 73

US-09-927-091-2
; Sequence 2, Application US/09927091
; Patent No. US20020119541A1
; GENERAL INFORMATION:
; APPLICANT: KILLARY, ANN
; APPLICANT: LOTY, STEVE
; APPLICANT: CHANDLER, DAMN
; TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1
; FILE REFERENCE: UTSC:651US
; CURRENT APPLICATION NUMBER: US/09/927,091
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/227,560
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 60/225,033
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 304
; TYPE: PRT
; ORGANISM: Human
US-09-927-091-2

Query Match 1.3%; Score 8; DB 9; Length 304;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LAPATAG 190
|||||
DB 283 LAPATAG 290

RESULT 74

US-10-424-599-185947
; Sequence 185947, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 185947
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138923C.1.pap
US-10-424-599-185947

Query Match 1.3%; Score 8; DB 15; Length 314;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 IEELKSER 254
|||
Db 158 IEELKSER 165

RESULT 75
US-10-416-387-2
; Sequence 2, Application US/10416387
; Publication No. US20040053347A1
; GENERAL INFORMATION:
; APPLICANT: Shiehatcar, Ramlin
; APPLICANT: The Wistar Institute
; TITLE OF INVENTION: BRAF35 Protein and BRCA2/BRAF35 Complex and Methods of
; TITLE OF INVENTION: Use
; FILE REFERENCE: WSTR-0011
; CURRENT APPLICATION NUMBER: US/10/416,387
; CURRENT FILING DATE: 2003-05-09
; PRIOR APPLICATION NUMBER: 60/247,618
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-416-387-2

Query Match 1.3%; Score 8; DB 15; Length 315;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAGG 17
|||
Db 10 AAAAPAGG 17

Search completed: March 23, 2005, 16:04:47
Job time : 144 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2005, 15:34:00 ; Search time 167 Seconds

(without alignments)
1421.981 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 614
Sequence: 1 MTRSAVFRAAAPAGNPE.....RSRCRLRIQKQTPHRTCRK 614

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : A.GeneSeq_16Dec04:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	614	100.0	614	6	ABR58310
2	278	45.3	281	3	AAB58317
3	243	39.6	269	3	AAB42665
4	153	24.9	153	8	ADQ66392
5	139	22.6	380	7	ADFS6888
6	116	18.9	189	7	ADCO6841
7	116	18.9	192	7	ADCO6839
8	48	7.8	60	8	ADP84563
9	47	7.7	203	7	ADC06840
10	25	4.1	195	4	AAU30711
11	25	4.1	195	4	ADP60190
12	10	1.6	1315	4	ABG09961
13	10	1.6	1353	4	ADA27376
14	10	1.6	1353	8	ADS10793
15	10	1.6	1390	5	ABG96292
16	10	1.6	1390	8	ADQ19160
17	9	1.5	85	4	AAO11030
18	9	1.5	224	4	ABO59348
19	9	1.5	242	8	ABO59348
20	9	1.5	288	4	AAAM80259
21	9	1.5	323	6	ABR58655
22	9	1.5	323	8	ADR14446
23	9	1.5	323	8	ABM82292
24	9	1.5	449	8	ADQ67687
25	9	1.5	548	4	ADG27726
26	9	1.5	553	6	AAO16423
27	9	1.5	652	7	ADBE5210
28	9	1.3	10	4	AAAG83786
29	9	1.3	10	4	AAAG83856
30	9	1.3	10	4	AAAG84004
31	9	1.3	10	4	AAAG84110
32	9	1.3	10	4	AAAG84006
33	9	1.3	15	2	AAAR75542
34	9	1.3	15	5	ABP52300
35	9	1.3	15	7	ADP41646
36	9	1.3	17	5	ABP52302
37	9	1.3	31	8	ADN11208
38	9	1.3	31	8	ADN11225
39	9	1.3	31	8	ADN11207
40	9	1.3	31	8	ADN11226
41	9	1.3	64	3	AAAG02201
42	9	1.3	79	5	ABP10891
43	9	1.3	80	3	ABR58244
44	9	1.3	83	2	AAAM19092
45	9	1.3	83	2	AAAY23318
46	9	1.3	98	2	AAW66330
47	9	1.3	109	3	AAAG34350
48	9	1.3	109	6	AAAG35343
49	9	1.3	112	4	AAAG3554
50	9	1.3	112	8	ADN24575
51	9	1.3	115	3	AAAG41132
52	9	1.3	123	3	AAAG01944
53	9	1.3	140	4	ABG19967
54	9	1.3	142	3	AAAY3408
55	9	1.3	143	3	AAAG41131
56	9	1.3	146	5	ABP43613
57	9	1.3	146	5	ABP43613
58	9	1.3	146	8	ABM81840
59	9	1.3	167	3	AAAG32017
60	9	1.3	169	7	ABO61854
61	9	1.3	171	3	AAAG54374
62	9	1.3	171	3	AAAG42727
63	9	1.3	171	3	AAAG3183
64	9	1.3	176	6	ABU06773
65	9	1.3	180	2	AAAT5508
66	9	1.3	184	7	ABO73997
67	9	1.3	187	3	AAAG32016
68	9	1.3	191	4	ABR21687
69	9	1.3	191	4	AAAG69485
70	9	1.3	191	4	AAAG57094
71	9	1.3	191	4	AAAG5009
72	9	1.3	191	4	ADG27706
73	9	1.3	202	8	ADCO6567
74	9	1.3	208	8	ABO60152
75	9	1.3	216	4	ABR67341
76	9	1.3	223	2	ABO78137
77	9	1.3	232	2	AAV08856
78	9	1.3	232	2	AAV08769
79	9	1.3	247	4	AAAM80040
80	9	1.3	249	7	ADH44202
81	9	1.3	252	8	ADH44202
82	9	1.3	252	8	ADH44201
83	9	1.3	261	7	ADH44199
84	9	1.3	262	7	ADH44200
85	9	1.3	263	8	ADH44200
86	9	1.3	265	8	ADH44200
87	9	1.3	266	5	ABG79694
88	9	1.3	274	4	AAE02490
89	9	1.3	289	3	AAAG42726
90	9	1.3	289	3	AAAG4373
91	9	1.3	289	3	AAAG4373
92	9	1.3	292	5	ABG73657
93	9	1.3	292	5	ABG73657
94	9	1.3	292	8	ADH44202
95	9	1.3	292	8	ADH44202
96	9	1.3	304	5	AAU78658
97	9	1.3	305	3	AAAG32015
98	9	1.3	309	3	AAAG54372
99	9	1.3	309	3	AAAG54372
100	9	1.3	309	3	AAAG54372

99 8 1.3 312 3 AAG23182 Aag23182 Arabidops
100 8 1.3 312 3 AAG42725 Aag42725 Arabidops

ALIGNMENTS

RESULT 1

ABR58310
ID ABR58310 standard; protein; 614 AA.

AC ABR58310;

DT 08-JUL-2003 (first entry)

DE BCUI041 protein #SEQ ID 66.

XX Breast cancer; cytostatic; gene therapy; antisense therapy; regulated;
KM drug discovery; clinical medicine; forensic medicine; chromosome 7q21.1.

XX Homo sapiens.

XX WO2003029421-A2.

XX 10-APR-2003.

XX 02-OCT-2002; 2002WO-US031287.

XX 03-OCT-2001; 2001US-0326526P.

XX 14-MAY-2002; 2002US-00144194.

XX (ORIG-) ORIGENE TECHNOLOGIES INC.

XX Sun Z, Li X, Fan W, Kovacs KF, Jay G;

XX WPI; 2003-381623/36.

XX N-PSDB; ACC72044.

XX New isolated human differentially-regulated breast cancer polynucleotide
PT and polypeptide, useful for diagnosing, staging, prognosticating,
PT preventing and/or treating diseases and conditions relating to breast
XX cancer.

XX Claim 6; SEQ ID NO 66; 127pp + Sequence listing; English.

XX The invention relates to isolated polynucleotides which are
XX differentially-regulated in breast cancer. The methods and compositions
XX of the present invention are useful for detecting, diagnosing, staging,
XX monitoring, prognosticating, preventing and/or treating diseases and
XX conditions relating to breast cancer, and may be used in gene therapy or
XX antisense therapy. They can also be used in research, drug discovery,
XX clinical medicine and forensic medicine. Sequences given in records
XX ABR58278-ABR58323 and ABR58346-ABR58362 represent polypeptides encoded by
XX polynucleotides of the invention that are differentially-regulated in
XX breast cancer. NOTE: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 614 AA;

Query Match 100.0%; Score 614; DB 6; Length 614;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 614; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTRSAVFKAAPAGNPEORLDYERAAALGGPEDESGAABHFLPBRHKLKEGPPILA 60

DB 1 MTRSAVFKAAPAGNPEORLDYERAAALGGPEDESGAABHFLPBRHKLKEGPPILA 60

QY 61 SSGGSGPAPBPACGGKGRGILLPAGAAPGQESWGSVPLPCPPATKQAGIGEPAA 120

DB 61 SSGGSGPAPBPACGGKGRGILLPAGAAPGQESWGSVPLPCPPATKQAGIGEPAA 120

QY 121 AGAGCSPRPKYQAVLPITQISLVAAAKETPMAAGDKGGAASPAATASDPAGPPPLP 180

DB 121 AGAGCSPRPKYQAVLPITQISLVAAAKETPMAAGDKGGAASPAATASDPAGPPPLP 180

QY 181 PPLAPATAGTLAASGRWKSMPKSPFGGGGSGAASSQAACLKQIILLQLDLIEGQOQUL 240

DB 181 PPLAPATAGTLAASGRWKSMPKSPFGGGGSGAASSQAACLKQIILLQLDLIEGQOQUL 240

QY 241 QAKKEKEIEELKSRDPLLARIRFMRMRMLVKNDKEKHKLFOGYETERESTELSEKI 300

DB 241 QAKKEKEIEELKSRDPLLARIRFMRMRMLVKNDKEKHKLFOGYETERESTELSEKI 300

QY 301 KLECPLESETSQLPPKPFSCGSGKHRSKPSFSTERTKTPVKKLAEPSKYTKTPK 360

DB 301 KLECPLESETSQLPPKPFSCGSGKHRSKPSFSTERTKTPVKKLAEPSKYTKTPK 360

QY 361 HSPIKEPCGSLSETVCKELRSQETPEKRSSVDTPPLSTPQKPSHPKKAFFSSEI 420

DB 361 HSPIKEPCGSLSETVCKELRSQETPEKRSSVDTPPLSTPQKPSHPKKAFFSSEI 420

QY 421 EDLPYSTTEMILCRWHOPPPSPPLPRESPPKEETVACLMPSVAGETSVLAVPSMD 480

DB 421 EDLPYSTTEMILCRWHOPPPSPPLPRESPPKEETVACLMPSVAGETSVLAVPSMD 480

QY 481 HSYEPRLRDNPSDLLENLDSVFSKHAKELEDEKRRKMDIQIREQRIQLRLMYK 540

DB 481 HSYEPRLRDNPSDLLENLDSVFSKHAKELEDEKRRKMDIQIREQRIQLRLMYK 540

QY 541 KKGIOSEBEVTSFPPEPDVSLMTPLPVVAFGRPLPKLTPQNFEIPWIDERSRCL 600

DB 541 KKGIOSEBEVTSFPPEPDVSLMTPLPVVAFGRPLPKLTPQNFEIPWIDERSRCL 600

QY 601 EIQKQTPHRTCRK 614

DB 601 EIQKQTPHRTCRK 614

RESULT 2

AAB58317
ID AAB58317 standard; protein; 281 AA.

AC AAB58317;

XX 14-MAR-2001 (first entry)

XX Lung cancer associated polypeptide sequence SEQ ID 655.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;

XX cardiotoxic; immunomodulatory; muscular active; vulnery;

XX gastrointestinal; nephrotoxic; antiinfective; gynecological;

XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;

XX proliferative disorder; wound healing; infectious disease.

XX Homo sapiens.

XX WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US005918.

XX 12-MAR-1999; 99US-0124270P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Ruben SM.

XX WPI; 2000-587514/55.

XX N-PSDB; AAF18193.

XX Lung cancer associated gene sequences, referred to as lung cancer

XX antigens, useful for treatment, prevention, and diagnosis of disorders

XX such as lung cancer.

XX Claim 11; Page 1153-1154; 1425bp; English.
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnary; gastrointestinal
 CC general; nephrotoxic; antineoplastic; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the protein
 CC or polynucleotide sequences. The lung cancer associated polynucleotide
 CC sequences may be used for detection of lung cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The proteins may be used to treat disorders such as
 CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
 CC cardiovascular, renal, and proliferative disorders. The proteins may also
 CC be used in the treatment of wounds and infectious diseases.
 CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
 CC used in the course of the invention for the identification and
 CC characterization of the polynucleotide and protein sequences

XX Sequence 281 AA:

Query Match 45.3%; Score 278; DB 3; Length 281;
 Best Local Similarity 100.0%; Pred. No. 5.2e-248;
 Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 PPLAPATATGTTAASGRWMSMKSPGCGGSGGASQAACIQLIILDLIEQQOOL 240
 DB 1 PPLAPATATGTTAASGRWMSMKSPGCGGSGGASQAACIQLIILDLIEQQOOL 60
 QY 241 QAKKEIEELKSRDITLARIEMERRMOLVKDNEKERRKLFGYETEEREETLESEKI 300
 DB 61 QAKKEIEELKSRDITLARIEMERRMOLVKDNEKERRKLFGYETEEREETLESEKI 120
 QY 301 KLECOPELSETSQTLPPKPFSCGRSGKGRKSPFGSTERKTVPKLAPEFSKYKTKTPK 360
 DB 121 KLECOPELSETSQTLPPKPFSCGRSGKGRKSPFGSTERKTVPKLAPEFSKYKTKTPK 180
 QY 361 HSEIKPEPCGSLSEYCKRELBQSETPKRRSSVDPPLSTQKGPSTHPKKAFFSEI 420
 DB 181 HSEIKPEPCGSLSEYCKRELBQSETPKRRSSVDPPLSTQKGPSTHPKKAFFSEI 240
 QY 421 EDLPYSTTEMYLCRMHQPSPPLPLRESPPKKEETVA 458
 DB 241 EDLPYSTTEMYLCRMHQPSPPLPLRESPPKKEETVA 278

RESULT 3

AAB42665
 ID AAB42665 standard; protein; 269 AA.

XX AAB42665;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF2429 polypeptide sequence SEQ ID NO:4858.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 XX vulnary; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
 XX anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
 XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 XX hypotensive; dermatological; immunosuppressive; antineoplastic;
 XX antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 XX antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
 XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
 XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 XX cholesterol ester storage; systemic lupus erythematosus; infection;
 XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
 XX thrombosis; contraceptive.

OS Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000MO-US008621.

XX 31-MAR-1999; 99US-0127607P.

XX 02-APR-1999; 99US-0127636P.

XX 05-APR-1999; 99US-0127728P.

XX 30-MAR-2000; 2000US-00540763.

XX (CURA-) CURAGEN CORP.

XX Shimkete RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAC76874.

XX Claim 11; Page 4046; 5507bp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
 CC anticonvulsant; antirheumatic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antineoplastic; antibacterial;
 CC antiviral; antifungal; antineumatic; antihypertensive; antineoplastic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancer, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 269 AA:

Query Match 39.6%; Score 243; DB 3; Length 269;
 Best Local Similarity 100.0%; Pred. No. 1.1e-215;
 Matches 243; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 SSQAACIKOILLIQLDLIEQQOOLQAKKEIEELKSRDITLARIEMERRMOLVKDN 275
 DB 24 SSQAACIKOILLIQLDLIEQQOOLQAKKEIEELKSRDITLARIEMERRMOLVKDN 83
 QY 276 EKERHKLFGYETEEREETLESEKIKLECOPELSETSQTLPPKPFSCGRSGKGRKSPF 335
 DB 84 EKERHKLFGYETEEREETLESEKIKLECOPELSETSQTLPPKPFSCGRSGKGRKSPF 143
 QY 336 GSTERKTPVKLAPEFSKYKTKTPKSPKKEECSLSSEVCKRELRSQSTPEKRRSSVD 395
 DB 144 GSTERKTPVKLAPEFSKYKTKTPKSPKKEECSLSSEVCKRELRSQSTPEKRRSSVD 203
 QY 396 TPRRLSTPOKGPSTHPKKAFFSEIDLPYSTTEMYLCRMHQPSPPLPLRESPPKKEE 455
 DB 204 TPRRLSTPOKGPSTHPKKAFFSEIDLPYSTTEMYLCRMHQPSPPLPLRESPPKKEE 263
 QY 456 TVA 458
 DB 264 TVA 266

RESULT 4
ADQ66392
ID ADQ66392 standard; protein; 153 AA.
XX
AC ADQ66392;
XX
DT 07-OCT-2004 (first entry)
XX
DE Novel human protein sequence #1365.
XX
OS osteoparitic; neuroprotective; nootropic; antiparkinsonian; cyrostatic;
KW gene therapy; diagnostic marker; morbid state; osteoporosis;
KW neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW cancer.
XX
OS Homo sapiens.
XX
PN EP1440981-A2.
XX
PD 28-JUL-2004.
XX
PF 21-JAN-2004; 2004EP-00001196.
XX
PR 21-JAN-2003; 2003JP-00102206.
XX 09-MAY-2003; 2003JP-00131392.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S,
PI Yamamoto J, Isono Y, Nagai K, Irie R;
XX
DR MPI; 2004-535376/52.
XX
DR N-PSDB; ADQ64204.
XX
PT Novel 2495 CDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS Claim 1; SEQ ID NO 3553; 2449pp; English.
XX
CC The invention relates to 2495 novel polynucleotides (I) and their encoded
CC polypeptides, sequences hybridizing to these nucleotides, sequences
CC encoding partial polypeptides and sequences having 70% or 90% identity to
CC the nucleotide and protein sequences. The nucleotides and polypeptides
CC are useful as diagnostic markers or therapeutic target for the diseases
CC or morbid states. They are also useful for treating osteoporosis,
CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC dementia and various cancers. This sequence corresponds to a protein
CC sequence of the invention.
XX
SQ Sequence 153 AA;
XX
Query Match 24.9%; Score 153; DB 8; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.1e-132;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 462 MESSVAGETSVLAVPMSRDHSEVPLDPNPSDLLLENLSDVFSKRAKLELDEKRRKRW 521
DB 1 MESSVAGETSVLAVPMSRDHSEVPLDPNPSDLLLENLSDVFSKRAKLELDEKRRKRW 60
XX
QY 522 IQRIRQRILQRILQRLMYKKKGIOESBPEVTSFPPEDDVESLMTTPFLPVVAFGRPLPK 581
DB 61 IQRIRQRILQRILQRLMYKKKGIOESBPEVTSFPPEDDVESLMTTPFLPVVAFGRPLPK 120
XX
QY 582 LTFQNFELPWLDERSRCLLEIOKKQTPHRTCRK 614
DB 121 LTFQNFELPWLDERSRCLLEIOKKQTPHRTCRK 153
XX
RESULT 5
ADFS58688
ID ADFS58688 standard; protein; 380 AA.
XX

AC ADFS58688;
XX
DT 12-FEB-2004 (first entry)
XX
DE Human polypeptide sequence SEQ ID NO:1096.
XX
KW biological activity; genetic engineering; hybridisation probe; oligomer;
KW primer; chromosome mapping; gene mapping; recombinant protein production;
KW human.
XX
OS Homo sapiens.
XX
PN WO2003080795-A2.
XX
PD 02-OCT-2003.
XX
PF 09-AUG-2002; 2002WO-US025485.
XX
PR 09-AUG-2001; 2001US-0311261P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Yang Y, Wang Z, Weng G, Ma Y;
XX
DR MPI; 2003-876918/81.
XX
DR N-PSDB; ADFS7668.
XX
PS Claim 20; SEQ ID NO 1096; 571pp; English.
XX
CC The present sequence represents a polypeptide (II) with biological
CC activity, which is encoded by an isolated polynucleotide sequence (I)
CC from the present invention. Also described: (1) a vector comprising (I);
CC (2) an expression vector comprising (I); (3) a host cell genetically
CC engineered to comprise (I) which is operatively associated with a
CC regulatory sequence that modulates expression of (I) in the host cell;
CC (4) a polypeptide (II) encoded by (I); (5) a composition comprising the
CC polypeptide of (4) and a carrier; (6) an antibody directed against the
CC polypeptide of (4); (7) detecting (I) or the polypeptide of (4) in a
CC sample; (8) identifying a compound that binds to the polypeptide of (4);
CC (9) producing the polypeptide of (4); and (10) a collection of
CC polynucleotides comprising at least one of the polynucleotide sequences
CC (I). The polynucleotides (I) can be used as hybridisation probes,
CC oligomers or primers, for chromosome or gene mapping, for the recombinant
CC production of proteins, and for generating antisense DNA or RNA.
XX
SQ Sequence 380 AA;
XX
Query Match 22.6%; Score 139; DB 7; Length 380;
Best Local Similarity 100.0%; Pred. No. 2.1e-119;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 476 PSWRDHSVEPLDPNPSDLLLENLSDVFSKRAKLELDEKRRKWDIQIRQRILQRIQ 535
DB 242 PSWRDHSVEPLDPNPSDLLLENLSDVFSKRAKLELDEKRRKWDIQIRQRILQRIQ 301
XX
QY 536 LRMVKKKGIOESBPEVTSFPPEDDVESLMTTPFLPVVAFGRPLKLTQNFELPWLDER 595
DB 302 LRMVKKKGIOESBPEVTSFPPEDDVESLMTTPFLPVVAFGRPLKLTQNFELPWLDER 361
XX
QY 596 SRCRLLEIOKKQTPHRTCRK 614
DB 362 SRCRLLEIOKKQTPHRTCRK 380
XX
RESULT 6
ADCC06841
ID ADCC06841 standard; protein; 189 AA.
AC ADCC06841;
XX

XX 18-DEC-2003 (first entry)
XX
XX Human up-regulated breast cancer transcription factor truncated protein.
DE
XX cytostatic; prostate cancer; breast; gene therapy; transgenic; human;
KW up-regulated breast cancer transcription factor; urd-ctf;
KM chromosome 17q21.1.
XX
OS Homo sapiens.
XX
XX MO2003064599-A2.
XX
XX 07-AUG-2003.
XX
XX 24-JAN-2003; 2003MO-US001943.
XX
XX 25-JAN-2002; 2002US-00054935.
XX 14-FEB-2002; 2002US-0356130P.
XX 22-MAR-2002; 2002US-00102946.
XX 08-APR-2002; 2002US-00117229.
XX 14-MAY-2002; 2002US-00144198.
XX 19-JUL-2002; 2002US-00197824.
XX
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
XX Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;
XX
XX WPI; 2003-679495/64.
XX
XX New isolated polynucleotide related to cancer genes, useful for
PT detecting, diagnosing, staging, monitoring, prognosticating, preventing
PT or treating cancers, e.g. breast and prostate cancers.
PS Disclosure; Fig 21, 128pp; English.
XX
XX The invention relates to a novel isolated polynucleotide comprising a
CC differentially-regulated mammalian cancer gene. The polynucleotides of
CC the invention demonstrate cytostatic activity and are differentially
CC expressed in prostate cancer. The polynucleotide, polypeptides and
CC methods of the invention may be useful for detecting, diagnosing,
CC staging, monitoring, prognosticating, preventing or treating cancers,
CC particularly breast and prostate cancers. Furthermore, the invention may
CC be utilised during gene therapy procedures or in the production of
CC transgenic animals. The current sequence is that of the prostate cancer-
CC related protein of the invention.
XX
XX
SQ Sequence 189 AA;
Query Match 18.9%; Score 116; DB 7; Length 189;
Best Local Similarity 100.0%; Pred. No. 2.1e-98;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 343 PVKKLAPEFSKVTKTTPKHSPIKEBPGSLSETVCKRELRSQETPEKPRSSVDTPPRLST 402
Db 71 PVKKLAPEFSKVTKTTPKHSPIKEBPGSLSETVCKRELRSQETPEKPRSSVDTPPRLST 130
QY 403 POKGPTHPPEKAFSSSEIEDLPYLSTTEMYLCRWHPDPPLPIRESSPKKEETVA 458
Db 131 POKGPTHPPEKAFSSSEIEDLPYLSTTEMYLCRWHPDPPLPIRESSPKKEETVA 186
RESULT 7
ADC06839
ID ADC06839 standard; protein; 192 AA.
XX
XX
XX ADC06839;
XX
XX 18-DEC-2003 (first entry)
XX
XX Human up-regulated breast cancer transcription factor protein.
XX
XX cytostatic; prostate cancer; breast; gene therapy; transgenic; human;

KW up-regulated breast cancer transcription factor; urd-ctf;
KM chromosome 17q21.1.
XX
XX Homo sapiens.
XX
XX MO2003064599-A2.
XX
XX 07-AUG-2003.
XX
XX 24-JAN-2003; 2003MO-US001943.
XX
XX 25-JAN-2002; 2002US-00054935.
XX 14-FEB-2002; 2002US-0356130P.
XX 22-MAR-2002; 2002US-00102946.
XX 08-APR-2002; 2002US-00117229.
XX 14-MAY-2002; 2002US-00144198.
XX 19-JUL-2002; 2002US-00197824.
XX
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
XX Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;
XX
XX WPI; 2003-679495/64.
XX
XX New isolated polynucleotide related to cancer genes, useful for
PT detecting, diagnosing, staging, monitoring, prognosticating, preventing
PT or treating cancers, e.g. breast and prostate cancers.
PS Claim 5, Fig 21, 128pp; English.
XX
XX The invention relates to a novel isolated polynucleotide comprising a
CC differentially-regulated mammalian cancer gene. The polynucleotides of
CC the invention demonstrate cytostatic activity and are differentially
CC expressed in prostate cancer. The polynucleotide, polypeptides and
CC methods of the invention may be useful for detecting, diagnosing,
CC staging, monitoring, prognosticating, preventing or treating cancers,
CC particularly breast and prostate cancers. Furthermore, the invention may
CC be utilised during gene therapy procedures or in the production of
CC transgenic animals. The current sequence is that of the prostate cancer-
CC related protein of the invention.
XX
XX
SQ Sequence 192 AA;
Query Match 18.9%; Score 116; DB 7; Length 192;
Best Local Similarity 100.0%; Pred. No. 2.2e-98;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 343 PVKKLAPEFSKVTKTTPKHSPIKEBPGSLSETVCKRELRSQETPEKPRSSVDTPPRLST 402
Db 75 PVKKLAPEFSKVTKTTPKHSPIKEBPGSLSETVCKRELRSQETPEKPRSSVDTPPRLST 134
QY 403 POKGPTHPPEKAFSSSEIEDLPYLSTTEMYLCRWHPDPPLPIRESSPKKEETVA 458
Db 135 POKGPTHPPEKAFSSSEIEDLPYLSTTEMYLCRWHPDPPLPIRESSPKKEETVA 190
RESULT 8
ADP84563
ID ADP84563 standard; protein; 60 AA.
XX
XX
XX ADP84563;
XX
XX 09-SEP-2004 (first entry)
XX
XX Human breast-specific protein #67.
XX
XX human; breast-specific protein; breast cancer.
XX
XX Homo sapiens.
XX
XX WO2004053077-A2.
XX
XX 24-JUN-2004.

CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukemias.
CC AAU29510-AAU33104 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX

SQ Sequence 195 AA;

Query Match 4.1%; Score 25; DB 4; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.9e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAGNPEORLDYERAAALGCP 34
DB 1 AAAAPAGNPEORLDYERAAALGCP 25

RESULT 11

ADP60190
ID ADF60190 standard; protein; 195 AA.

XX ADF60190;

DT 12-FEB-2004 (first entry)

DE Human contig polypeptide sequence SEQ ID NO:2557.

XX biological activity; genetic engineering; hybridisation probe; oligomer;
KM primer; chromosome mapping; gene mapping; recombinant protein production;
KM human.

XX Homo sapiens.

PN MO2003080795-A2.

PD 02-OCT-2003.

PF 09-AUG-2002; 2002MO-US025485.

PR 09-AUG-2001; 2001US-0311261P.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Yang Y, Wang Z, Wang G, Ma Y;

DR MPI: 2003-876918/81.

DR N-PSDB; ADF59738.

XX New polynucleotides, useful as hybridization probes, oligomers or
PT primers, for chromosome or gene mapping, for the recombinant production
PT of proteins, and for generating antisense DNA or RNA.

PS Example 3; SEQ ID NO 2557; 571bp; English.

XX The present invention describes isolated polynucleotide sequences (I),
CC which encode polypeptides (II) with biological activity. Also described:
CC (1) a vector comprising (I); (2) an expression vector comprising (I); (3)
CC a host cell genetically engineered to comprise (I) which is operatively
CC associated with a regulatory sequence that modulates expression of (I) in
CC the host cell; (4) a polypeptide (II) encoded by (I); (5) a composition
CC comprising the polypeptide of (4) and a carrier; (6) an antibody directed
CC against the polypeptide of (4); (7) detecting (I) or the polypeptide of
CC (4) in a sample; (8) identifying a compound that binds to the polypeptide
CC of (4); (9) producing the polypeptide of (4); and (10) a collection of
CC polynucleotides comprising at least one of the polynucleotide sequences
CC (I). The polynucleotides (I) can be used as hybridization probes,
CC oligomers or primers, for chromosome or gene mapping, for the recombinant
CC production of proteins, and for generating antisense DNA or RNA. The
CC present sequence represents a human contig polypeptide sequence, which is
CC used in an example from the present invention.

SQ Sequence 195 AA;

Query Match 4.1%; Score 25; DB 7; Length 195;

Best Local Similarity 100.0%; Pred. No. 2.9e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 AAAAPAGNPEORLDYERAAALGCP 34
DB 1 AAAAPAGNPEORLDYERAAALGCP 25

RESULT 12

ABG09961
ID ABG09961 standard; protein; 1315 AA.

XX ABG09961;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #9952.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN MO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR MPI: 2001-639362/73.

DR N-PSDB; AAS74148.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PS Claim 20; SEQ ID NO 40320; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 1315 AA;

Query Match 1.6%; Score 10; DB 4; Length 1315;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGASS 217
|||
DB 62 GGGGSGGASS 71

RESULT 13
ADA27376 standard; protein, 1353 AA.
XX
AC ADA27376;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human CAPS protein SEQ ID NO:1.
XX
KW neurodegenerative disease; Alzheimer's disease; human;
KW calcium-dependent activator protein for secretion; CAPS; neuronal;
KW neuroendocrine.
XX
OS Homo sapiens.
XX
PN WO2003069347-A2.
XX
PD 21-AUG-2003.
XX
PF 14-FEB-2003; 2003WO-EP001493.
XX
PR 14-FEB-2002; 2002EP-00003431.
XX
PR 14-FEB-2002; 2002US-0356155P.
XX
PA (EVOT-) EVOTEC NEUROSCIENCES GMBH.
XX
PI Hipfel R, Von Der Kammer H, Pohlner J;
XX
XX WPI; 2003-671683/63.
XX
DR Diagnosing, prognosing or evaluating a treatment for Alzheimer's disease,
XX PT comprises determining the level or activity of a transcription and/or
XX translation product of a gene coding for an activator protein for vesicle
XX secretion.
XX
PS Disclosure; Fig 4; 52pp; English.
XX
XX The present invention describes a method for diagnosing, prognosticating,
XX and monitoring the progression of, or evaluating a treatment for, a
XX neurodegenerative disease in a subject, or determining if a subject is at
XX increased risk of developing the disease. The method comprises
XX determining the level or activity of a transcription and/or translation
XX product of a gene coding for an activator protein for vesicle secretion,
XX and/or its fragment, derivative or variant, in a sample and comparing the
XX level and/or activity to a reference value representing a known disease
XX or health status. The method is useful in diagnosing, preventing or
XX treating neurodegenerative diseases, particularly Alzheimer's disease.
XX The present sequence represents a human calcium-dependent activator
XX protein for secretion (CAPS) which is a protein restricted in its
XX expression to neuronal and neuroendocrine cells, and is used in the
XX exemplification of the present invention.
XX
SQ Sequence 1353 AA;

Query Match 1.6%; Score 10; DB 7; Length 1353;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGASS 217
|||
DB 62 GGGGSGGASS 71

RESULT 14
ADS10793
ID ADS10793 standard; protein, 1353 AA.

XX
AC ADS10793;
XX
DT 16-DEC-2004 (first entry)
XX
DE Human therapeutic protein - SEQ ID NO 1030.
XX
KW antiinflammatory; neuroprotective; antianaemic; cyrostatic; vulnerary;
KW inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;
KW aplastic anaemia; cancer; wound healing; gene therapy.
XX
OS Homo sapiens.
XX
PN WO2004080148-A2.
XX
PD 23-SEP-2004.
XX
PF 30-SEP-2003; 2003WO-US030720.
XX
PR 02-OCT-2002; 2002US-0416186P.
XX
XX (NUVE-) NUVELO INC.
XX
PA
XX
PI Tang YT, Asundi V, Ren F, Zhang J, Wehrman T, Wang Z, Ma Y;
XX Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;
XX WPI; 2004-668857/65.
XX
DR N-PSDB; ADS10109.
XX
XX
XX
PT New polynucleotide, useful in preparing a composition for diagnosing or
PT treating inflammatory, neurodegenerative or stem cell disorders, e.g.,
PT aplastic anemia or cancer for promoting wound healing.
XX
XX Claim 20; SEQ ID NO 1030; 718pp; English.
XX
XX The invention relates to a novel isolated polynucleotide and the encoded
XX polypeptide. The molecules of the invention demonstrate antiinflammatory,
XX neuroprotective, antianaemic, cyrostatic and vulnerary activities and may
XX be useful in preparing a composition for diagnosing or treating
XX inflammatory, haematopoietic, immune, neurodegenerative or stem cell
XX disorders, such as aplastic anaemia or cancer, as well as for promoting
XX wound healing. The molecules may also be utilised during gene therapy
XX procedures. The current sequence is that of a human therapeutic protein
XX of the invention. The current sequence is not shown explicitly within the
XX specification but can be accessed from the WIPO web-site.
XX
SQ Sequence 1353 AA;

Query Match 1.6%; Score 10; DB 8; Length 1353;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGASS 217
|||
DB 62 GGGGSGGASS 71

RESULT 15
ABG96292
ID ABG96292 standard; protein, 1390 AA.
XX
AC ABG96292;
XX
DT 11-DEC-2002 (first entry)
XX
DE Human ovarian cancer marker M432.
XX
KW Human; ovarian cancer; marker; cancer; familial history; brain disorder;
KW central nervous system disorder; bacterial meningitis; viral meningitis;
KW Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
KW brain herniation; inflammation; encephalitis; testicular disorder;
KW nondercutaneous granulomatous orchitis; connective tissue disorder;
KW heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;

KM	histological type; carcinogenic; ovarian cancer marker.
OS	Homo sapiens.
XX	
PN	WO200271928-A2.
PD	19-SEP-2002.
XX	
PF	14-MAR-2002; 2002WO-US007826.
XX	
PR	14-MAR-2001; 2001US-0276025P.
PR	14-MAR-2001; 2001US-0276026P.
PR	10-AUG-2001; 2001US-0311732P.
PR	19-SEP-2001; 2001US-0323588P.
PR	26-SEP-2001; 2001US-0324967P.
PR	26-SEP-2001; 2001US-0325102P.
PR	26-SEP-2001; 2001US-0325149P.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
PI	Monahan JE, Gannavarapu M, Hoersch S, Kamathar S, Kovatis SG;
PI	Meyers RE, Morrissey MP, Olandt PJ, Sen A, Viehy PO, Mills GB;
PI	Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;
DR	WPI; 2002-723277/78.
DR	N-PDSB; ABST6384.
XX	
PT	Assessing whether a patient is afflicted with ovarian cancer, useful in
PT	assessing the stage or progression of the disease, comprises comparing
PT	the expression level of a cancer marker in a sample from a patient and
PT	from a non cancer patient.
XX	
PS	Disclosure; Page 148-151; 481pp; English.
XX	
CC	The present invention relates to a new method for assessing whether a
CC	patient is afflicted with ovarian cancer. The method involves comparing
CC	the expression level of a marker in a patient sample and the normal level
CC	of expression of the marker in a control non-ovarian cancer sample, where
CC	the marker is selected from 363 cancer markers described in the
CC	specification. The method of the invention is useful in diagnosing or
CC	characterising cancer, in detecting the presence of cancer as early as
CC	possible, and the recurrence of ovarian cancer. The method may also be of
CC	particular use with patients having an enhanced risk of developing
CC	ovarian cancer (e.g. patients having a familial history of ovarian
CC	cancer). The cancer markers may be used in the management and treatment
CC	of e.g. brain and central nervous system disorders (e.g. bacterial and
CC	viral meningitis, Alzheimer's disease or Parkinson's disease), brain
CC	disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),
CC	inflammations (e.g. bacterial or viral meningitis or encephalitis),
CC	testicular disorders (e.g. nontuberculous granulomatous orchitis),
CC	connective tissue disorders, or heart disorders (e.g. ischaemic heart
CC	disease or atherosclerosis). The compositions and methods may also be
CC	used in assessing the histological type of neoplasm associated with
CC	ovarian cancer, monitoring the progression of ovarian cancer, determining
CC	whether ovarian cancer has metastasized or is likely to metastasize,
CC	selecting a composition for inhibiting ovarian cancer, assessing the
CC	ovarian carcinogenic potential of a compound, or inhibiting ovarian
CC	cancer or at risk of developing ovarian cancer. The present amino acid
CC	sequence represents one of the ovarian cancer markers described in the
CC	invention
XX	
SQ	Sequence 1390 AA;
Query Match	1.6%; Score 10; DB 5; Length 1390;
Best Local Similarity	100.0%; Pred. No. 12;
Matches	10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	208 GGCGGSGASS 217
Dd	99 GGCGGSGASS 108

AD01	9160	standard; protein; 1390 AA.
ID	AD019160	
XX		
AC	AD019160;	
XX		
DT	26-AUG-2004	(first entry)
XX		
DE	Human soft tissue sarcoma-upregulated protein - SEQ ID 1979.	
XX		
KW	soft tissue sarcoma; cytosstatic; gene therapy; vaccine; screening; human.	
XX		
OS	Homo sapiens.	
XX		
PN	WO2004048938-A2.	
XX		
PD	10-JUN-2004.	
XX		
PF	26-NOV-2003; 2003WO-US038193.	
XX		
PR	26-NOV-2002; 2002US-0429739P.	
XX		
PA	(PROT-) PROTEIN DESIGN LABS INC.	
XX		
PI	Aziz N, Ginsburg WM, Zlotnik A;	
XX		
DR	WPI; 2004-441208/41.	
XX		
PT	Early detection of soft tissue sarcoma comprises determining expression	
PT	of a gene in a first soft tissue sample and a normal soft tissue sample	
PT	and comparing the gene expression, also useful in treating soft tissue	
PT	sarcoma.	
XX		
PS	Example 2; SEQ ID NO 1979; 210pp; English.	
XX		
CC	The invention relates to a novel method for detecting soft tissue sarcoma	
CC	which comprises obtaining a first soft tissue sample from an individual	
CC	and a normal soft tissue sample from the same or different individual,	
CC	determining the expression of a gene in both samples and comparing the	
CC	expression of the gene in both soft tissue samples, where a higher level	
CC	of protein expression in the first soft tissue sample indicates the	
CC	presence of soft tissue sarcoma. The method of the invention has	
CC	cytostatic applications and may be useful for detecting soft tissue	
CC	sarcoma, possibly via gene therapy or vaccine production. The nucleic	
CC	acid sequences may be useful in diagnostic and screening applications.	
CC	The current sequence is that of a human soft tissue sarcoma-upregulated	
CC	protein of the invention. The current sequence is not shown within the	
CC	specification per se but was submitted in CD format by the inventor.	
XX		
SQ	Sequence 1390 AA;	
Query Match	1.6%; Score 10; DB 8; Length 1390;	
Best Local Similarity	100.0%; Pred. No. 12;	
Matches 10; Conservative	0; Mismatches	0; Indels
		0; Gaps
		0;
Cy	208 GGCGGSGAGSS 217	
Db	99 GGCGGSGAGSS 108	
RESULT 17		
AAO11030		
ID	AAO11030 standard; protein; 85 AA.	
XX		
AC	AAO11030;	
XX		
DT	06-NOV-2001 (first entry)	
XX		
DE	Human polypeptide SEQ ID NO 24922.	
XX		
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	
KW	tissue growth factor; immunomodulatory; cancer; Leukaemia;	
KW	nervous system disorders; arthritis; inflammation.	

XX Homo sapiens.
OS
XX WO200164835-A2.
XX
XX 07-SEP-2001.
PD
XX
XX 26-FEB-2001; 2001WO-US004927.
PF
XX 28-FEB-2000; 2000US-00515126.
XX PR 18-MAY-2000; 2000US-00577409.
XX
XX (HYSEQ-) HYSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT;
PI
XX WPI; 2001-514838/56.
XX
XX N-PSDB; AAI90961.
DR
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
PT
XX
XX Claim 20; SEQ ID NO 24922; 1399bp + Sequence listing; English.
PS
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 85 AA;

Query Match 1.5%; Score 9; DB 4; Length 85;
Best Local Similarity 100.0%; Pred. No. 8; 7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 GPPPLPLPG 179
Db 20 GPPPLPLPG 28

RESULT 18
AB63455
ID AB63455 standard; protein; 224 AA.
XX
XX ABB63455;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 17157.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US009231.
XX PF
XX 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX

PA (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX N-PSDB; ABL07558.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
PT
XX
XX Disclosure; SEQ ID NO 17157; 21bp + Sequence listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 224 AA;

Query Match 1.5%; Score 9; DB 4; Length 224;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 OQOQOQOAK 243
Db 40 OQOQOQOAK 48

RESULT 19
ABO59348
ID ABO59348 standard; protein; 242 AA.
XX
XX ABO59348;
AC
XX
XX 29-JUL-2004 (first entry)
DT
XX
XX Human genome derived single exon protein #5582.
DE
XX
XX Human; gene expression; single exon probe; microarray;
KM alternative splicing event; genomic alteration.
XX
XX Homo sapiens.
OS
XX
XX US2003194704-A1.
PN
XX
XX 16-OCT-2003.
PD
XX
XX 03-APR-2002; 2002US-00029386.
XX PF
XX
XX 03-APR-2002; 2002US-00029386.
XX PR
XX
XX (PENN/) PENN S G.
PA (RANK/) RANK D R.
PA (HANZ/) HANZEL D K.
XX
XX Penn SG, Rank DR, Hanzel DK;
PI
XX
XX WPI; 2004-119264/12.
DR
XX
XX New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for
PT surveying tissues.
PT
XX
XX Claim 45; SEQ ID NO 32982; 80bp; English.
XX

CC The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule
CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above-mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC methods of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC human gene expression data by subscription, and a computer-readable
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above. The probe, methods and apparatus are useful in gene
CC expression analysis. The probes may be used as tools for surveying
CC tissues to detect the presence of expressed messages that contain their
CC specific exon, or in constructing genome-derived single exon microarrays.
CC In addition, the probes are used in identifying and characterising
CC alternative splicing events, in detecting and characterising gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe protein of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20030194704
CC
XX
SQ Sequence 242 AA;

Query Match 1.5%; Score 9; DB 8; Length 242;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGGAS 216
|||||||
DB 193 GGGGSGGAS 201

RESULT 20
AAM80259
ID AAM80259 standard; protein; 288 AA.

AC AAM80259;

DT 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 3905.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.

OS Homo sapiens.

PN WO200157190-A2.

PD 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US004098.

PR 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.

PA (HYSEQ-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejrtman T, Goodrich R;

XX WPI; 2001-476283/51.

DR N-PSDB; AAK53392.

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
in diagnosis and gene therapy.

PS Claim 20; Page 464; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK0302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity, and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM00020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
CC
XX
SQ Sequence 288 AA;

Query Match 1.5%; Score 9; DB 4; Length 288;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 GGGGSGGAS 217
|||||||
DB 72 GGGGSGGAS 80

RESULT 21
ABR58655
ID ABR58655 standard; protein; 323 AA.

AC ABR58655;

DT 09-JUL-2003 (first entry)

DE Human cancer related protein SEQ ID NO:312.

XX Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;
KW heart disease; atherosclerosis; endometriosis.

OS Homo sapiens.

PN WO2003025138-A2.

PD 27-MAR-2003.

PF 17-SEP-2002; 2002WO-US029560.

PR 17-SEP-2001; 2001US-0323469P.

PR 20-SEP-2001; 2001US-0323887P.

PR 13-NOV-2001; 2001US-0350666P.

PR 08-FEB-2002; 2002US-0355145P.

PR 08-FEB-2002; 2002US-0355257P.

PR 12-APR-2002; 2002US-0372246P.

PA (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 PI Afar D, Aziz N, Gish KC, Hwezi PA, Mack DH, Wilson KE;
 PI Zlounik A;
 XX
 DR WPI; 2003-354600/33.
 DR N-PSDB; ACC72806.
 XX
 PT New genes that are up-regulated or down-regulated in cancers, useful as
 PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as
 PT therapeutic targets for screening drugs for treating these diseases.
 XX
 PS Claim 12; Page 755; 767pp; English.
 XX
 CC The present invention describes an isolated nucleic acid molecule, which
 CC comprises the sequence of any of the genes that are up-regulated or down-
 CC regulated in specific cancers (e.g. about 1031 genes up-regulated in
 CC acute lymphocytic leukemia). ACC72841 to ACC72860 represent cancer
 CC related gene nucleotide sequences which encode the proteins given in
 CC ABR58521 to ABR58709. Also described: (1) determining the presence or
 CC absence of a pathological cell in a patient; (2) an expression vector
 CC comprising a nucleic acid molecule described above; (3) a host cell
 CC comprising the vector; (4) an isolated polypeptide, which is encoded by
 CC the nucleic acid; (5) an antibody that specifically binds the polypeptide
 CC of (4); (6) specifically targeting a compound to a pathological cell in a
 CC patient by administering to the patient the antibody above; and (7) a
 CC drug screening assay. The nucleic acid is useful as diagnostic markers or
 CC therapeutic targets. In particular, the nucleic acid is useful for
 CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,
 CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,
 CC pancreas, prostate, skin and uterus), wounds, ischemia, heart diseases,
 CC atherosclerosis and endometriosis. The nucleic acid is also useful in
 CC drug screening, particularly for identifying agents for treating these
 CC pathologies
 CC
 SQ Sequence 323 AA;
 XX
 Query Match 1.5%; Score 9; DB 6; Length 323;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 208 GGGGGSGAS 216
 Db 274 GGGGGSGAS 282
 XX
 RESULT 22
 ADR14446
 ID ADR14446 standard; protein; 323 AA.
 AC ADR14446;
 XX
 DT 21-OCT-2004 (first entry)
 XX
 DE Human NF-kappaB pathway-associated protein SeqID447.
 XX
 XX NF-kappaB pathway; antiinflammatory; cytosolic; hepatotropic; virucide;
 KM antiarthritic; antirheumatic; gastrointestinal-gen; antiasthmatic;
 KM antiarteriosclerotic; immunomodulator; cerebroprotective; vasotropic;
 KM immunosuppressive; vulnerrary; gene therapy; immune disorder;
 KM inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;
 KM hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour;
 KM hyper-igm syndrome; hypohidrotic ectodermal dysplasia;
 KM X-linked anhidrotic ectodermal dysplasia; immunodeficiency;
 KM viral infection; HIV-1; HTLV-1; hepatitis B; hepatitis C; EBV; influenza;
 KM viral replication; host cell survival; evasion of immune response;
 KM rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;
 KM atherosclerosis; cachexia; euthyroid sick syndrome; stroke; BAE;
 KM autoimmune disorder; hyper immune activity;
 KM aberrant acute phase response; hypercongenital condition; birth defect;
 KM necrotic lesion; wound; organ transplant rejection;
 KM aberrant signal transduction; proliferating disorder; cancer;
 KM HIV propagation; human.

XX
 OS Homo sapiens.
 XX
 PN WO2004065577-A2.
 XX
 PD 05-AUG-2004.
 XX
 PF 13-JAN-2004; 2004WO-US000798.
 XX
 PR 14-JAN-2003; 2003US-0440068P.
 PR 12-MAY-2003; 2003US-0469757P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Nadler SG, Neubauer MG, Feder JN, Carman J;
 PI N-PSDB; ADR14447.
 DR
 DR WPI; 2004-562168/54.
 XX
 PT New isolated polynucleotides and polypeptides associated with NF-kappaB
 PT pathway, useful for diagnosing, treating, or preventing disorders or
 PT diseases associated with NF-kappaB pathway.
 XX
 PS Claim 1; SEQ ID NO 447; 237pp; English.
 XX
 CC This invention relates to the novel association of protein sequences (and
 CC the genes which encode them) to the NF-kappaB pathway. The invention may
 CC be useful for the production of compounds with an antiinflammatory,
 CC cytosolic, hepatotropic, virucide, antiarthritic, antirheumatic,
 CC gastrointestinal-gen, antiasthmatic, antiarteriosclerotic,
 CC immunomodulator, cerebroprotective, vasotropic, immunosuppressive or
 CC vulnerrary activity or for gene therapy. The proteins and nucleotides are
 CC useful for diagnosing, preventing, treating, or ameliorating conditions
 CC or diseases associated with the NF-kappaB pathway. The condition is an
 CC immune disorder, an inflammatory disorder, an inflammatory disorder
 CC related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis,
 CC hepatic disorders, Hodgkin's lymphomas, haematopoietic tumours, hyper-igm
 CC syndromes, hypohidrotic ectodermal dysplasia, X-linked anhidrotic
 CC ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, HTLV-1,
 CC hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell
 CC survival, evasion of immune responses, rheumatoid arthritis, inflammatory
 CC bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick
 CC syndrome, stroke, BAE, autoimmune disorders, disorders related to hyper
 CC immune activity, disorders related to aberrant acute phase response,
 CC hypercongenital conditions, birth defects, necrotic lesions, wounds,
 CC organ transplant rejection, conditions related to organ transplant
 CC rejection, disorders related to aberrant signal transduction,
 CC proliferating disorders, cancers and HIV propagation in cells infected
 CC with other viruses. The present sequence is that of a human protein which
 CC is subject to the novel association with the NF-kappaB pathway of the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 CC
 SQ Sequence 323 AA;
 XX
 Query Match 1.5%; Score 9; DB 8; Length 323;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 208 GGGGGSGAS 216
 Db 274 GGGGGSGAS 282
 XX
 RESULT 23
 ABR82292
 ID ABR82292 standard; protein; 323 AA.
 AC ABR82292;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Tumour-associated antigenic target (TAT) polypeptide PRO83211, SEQ:5888.

XX WO200179254-A1.
PN
XX
XX 25-OCT-2001.
PD
XX
XX 16-APR-2001; 2001WO-US008655.
PF
XX
XX 18-APR-2000; 2000US-00552929.
PR 22-SEP-2000; 2000US-00668317.
PR 24-OCT-2000; 2000US-00695783.
PR 01-DEC-2000; 2000US-00728628.
PR 26-JAN-2001; 2001US-00770160.
PR 13-FEB-2001; 2001US-00783066.
PR 22-MAR-2001; 2001US-00816828.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang Y^T, Asundi V, Zhou P, Xue A^J, Ren F, Zhang J, Wang J;
PI Yang Y, Zhao QA, Goodrich RW, Liu C, Drmanac RT, Ma Y, Wang Z;
PI Wehman T;
XX
XX WPI; 2001-607699/69.
DR
XX
XX Novel polynucleotides and encoded polypeptides (protein factors,
PT including, e.g., cytokines, such as lymphokines, interferons, and
PT circulating soluble factors) useful for treating, e.g., Parkinson's,
PT Alzheimer's, HIV and cancer.
XX
XX Claim 20; SEQ ID NO 493; 153bp; English.
XX
XX This invention relates to a novel isolated DNA sequence and the mature
CC proteins encoded by them. The invention may be useful in the development
CC of compositions with antiparkinsonian, haemostatic, nootropic,
CC neuroprotective, osteopachic, anti-HIV, protozoacide, antifungal,
CC immunosuppressive, antineumatic, antiarthritic, antidiabetic,
CC antiallergic, antiinflammatory, anticoagulant or cytosstatic activities.
CC In addition, the sequences of the invention may be useful for gene
CC therapy. The invention may be useful for the development of treatments
CC for Parkinson's, Alzheimer's, thrombocytopaenia, osteoporosis,
CC osteoarthritis, infections (including HIV, Leishmania, malaria, and
CC various fungal infections), autoimmune disorders such as multiple
CC sclerosis, rheumatoid arthritis, and insulin dependent diabetes mellitus,
CC allergic reactions and conditions (for example food allergies, insect
CC allergies and allergic rhinitis), coagulation disorders including
CC haemophilia, and cancer. Note: The amino acid sequences given in table 6
CC (SeqID 439-584) may have in frame stop codons or possible
CC insertions/deletions as shown in the table. The sequences allocated Seq
CC IDs 1-438 are not provided (even by reference) in the specification.
XX
XX Sequence 548 AA;
SQ
Query Match 1.5%; Score 9; DB 4; Length 548;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 208 GGGGSGAS 216
DB 499 GGGGSGAS 507
RESULT 26
ID AAO16423
XX AAO16423 standard; protein; 553 AA.
XX
XX AAO16423;
AC
XX
XX 10-APR-2003 (first entry)
DT
XX
XX Human nucleic acid-associated protein (NAAP) - SEQ ID No 20.
DE
XX
XX Human, nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;

KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.
XX
XX Homo sapiens.
OS
XX
XX WO2003000864-A2.
PN
XX
XX 03-JAN-2003.
PD
XX
XX 20-JUN-2002; 2002WO-US021179.
PF
XX
XX 22-JUN-2001; 2001US-0300518P.
PR 29-JUN-2001; 2001US-0301787P.
PR 29-JUN-2001; 2001US-0301792P.
PR 29-JUN-2001; 2001US-0301892P.
PR 29-JUN-2001; 2001US-0301893P.
PR 06-JUL-2001; 2001US-0303405P.
PR 06-JUL-2001; 2001US-0303442P.
PR 15-MAR-2002; 2002US-0364438P.
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX
XX Gandhi AR, Swarnakar A, Hafalia A^JA, Warren BA, Emerling BM;
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Porsythe IJ;
PI Rankumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;
PI Borowsky ML, Yao MG, Walla NK, Bandman O, Lal PG, Becha SD, Lee ST;
PI Richardson TW, Elliott VS, Luo W, Tang YT, Zebardjadian Y, Lu Y;
XX
XX WPI; 2003-201420/19.
DR
XX
XX N-PSDB; AALS1573.
DR
XX
XX New nucleic acid-associated proteins and polynucleotides, useful for
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS).
XX
XX Claim 1, Page 246-247, 312pp; English.
XX
XX The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of NAAP, such as: cell proliferative
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC the invention are useful for creating transgenic animals to model human
CC disease. The present amino acid sequence represents a human nucleic acid-
CC associated protein of the invention
XX
XX Sequence 553 AA;
SQ
Query Match 1.5%; Score 9; DB 6; Length 553;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 209 GGGGSGASS 217
DB 27 GGGGSGASS 35
RESULT 27
ID ADB65210
XX ADB65210 standard; protein; 652 AA.
XX
XX ADB65210;
AC
XX
XX 04-DEC-2003 (first entry)
DT
XX
XX Human protein encoded by clone TEST110000190.
DE
XX

KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
 KW cell regeneration; membrane protein; signal transduction-related protein;
 KW transcription-related protein; osteoporosis; neurological disease;
 KW cancer; tumour.
 XX
 OS Homo sapiens.
 XX
 PN EP1308459-A2.
 XX
 PD 07-MAY-2003.
 XX
 PF 28-MAR-2002; 2002EP-00007401.
 XX
 PR 05-NOV-2001; 2001JP-00379298.
 PR 25-JAN-2002; 2002US-00350978.
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX
 DR WPI; 2003-450961/43.
 DR N-PSDB; ADB63240.
 XX
 PT New polynucleotides and polypeptides, useful for developing a diagnostic
 PT marker or medicines for regulation of their expression and activity, or
 PT as targets of gene therapy.
 XX
 PS Claim 1; Page: 222pp; English.
 XX
 SQ The invention discloses a polynucleotide comprising a sequence selected
 CC from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or
 CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein, and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesising the polynucleotide, or as a probe
 CC for detecting the polynucleotide. The polynucleotides and encoded
 CC proteins are useful as pharmaceutical agents and many disease-related
 CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell
 CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC transcription-related proteins, disease-related proteins and genes
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
 CC the activity or expression of the encoded protein to treat diseases. The
 CC sequence data presented is a protein of the invention. Note: Some of the
 CC sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 CC
 XX
 SQ Sequence 652 AA;
 XX
 QY Query Match 1.5%; Score 9; DB 7; Length 652;
 Db Best Local Similarity 100.0%; Pred. No. 51;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 208 GGGGGSGAS 216
 Db 562 GGGGGSGAS 570
 XX
 RESULT 28
 ID AAG83786 standard; peptide; 10 AA.
 XX
 AC AAG83786;

XX
 DT 11-SEP-2001 (first entry)
 XX
 XX Arabidopsis thaliana peptide ligand #426.
 DE
 XX
 KW Plant; peptide pesticide; peptide herbicide; agricultural research.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200142279-A2.
 XX
 PD 14-JUN-2001.
 XX
 PF 13-DEC-2000; 2000WO-GH004781.
 XX
 PR 13-DEC-1999; 99GB-00029469.
 XX
 PA (PROT-) PROTEOM LTD.
 XX
 PI Roberts GW, Heal JR;
 XX
 DR WPI; 2001-381629/40.
 XX
 PT A set of peptide ligands for agricultural research and development of
 PT therapeutic agents comprise specific complementary peptides to proteins
 PT encoded by genes of plant genomes.
 XX
 PS Example 4; Page 90; 201pp; English.
 XX
 SQ The present invention relates to a set of peptide ligands consisting of
 CC specific complementary peptides to proteins encoded by genes of plant
 CC genomes. The present sequence is one such peptide from Arabidopsis
 CC thaliana. The peptides of the present invention are useful in an assay to
 CC identify a peptide, especially a peptide pesticide or herbicide. The
 CC peptides are also useful for tools for agricultural research and
 CC development
 CC
 XX
 SQ Sequence 10 AA;
 XX
 QY 208 GGGGGSGA 215
 Db 3 GGGGGSGA 10
 XX
 RESULT 29
 ID AAG83856 standard; peptide; 10 AA.
 XX
 AC AAG83856;
 XX
 DT 11-SEP-2001 (first entry)
 XX
 DE Arabidopsis thaliana peptide ligand #496.
 XX
 KW Plant; peptide pesticide; peptide herbicide; agricultural research.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200142279-A2.
 XX
 PD 14-JUN-2001.
 XX
 PF 13-DEC-2000; 2000WO-GH004781.
 XX
 PR 13-DEC-1999; 99GB-00029469.
 XX
 PA (PROT-) PROTEOM LTD.
 XX
 PI Roberts GW, Heal JR;

```
XX WPI; 2001-381629/40.
XX
XX A set of peptide ligands for agricultural research and development of
PT therapeutic agents comprise specific complementary peptides to proteins
PT encoded by genes of plant genomes.
XX
XX Example 4; Page 100; 201pp; English.
XX
XX The present invention relates to a set of peptide ligands consisting of
CC specific complementary peptides to proteins encoded by genes of plant
CC genomes. The present sequence is one such peptide from Arabidopsis
CC thaliana. The peptides of the present invention are useful in an assay to
CC identify a peptide, especially a peptide pesticide or herbicide. The
CC peptides are also useful for tools for agricultural research and
CC development
XX
XX Sequence 10 AA;
SQ

Query Match      1.3%; Score 8; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      208 GGGGSGSGA 215
Db      1 GGGGSGSGA 8

RESULT 30
AAG84004
ID AAG84004 standard; peptide; 10 AA.
AC
AC AAG84004;
XX
XX 11-SEP-2001 (first entry)
XX
XX Arabidopsis thaliana peptide ligand #644.
DE
XX Plant; peptide pesticide; peptide herbicide; agricultural research.
XX
XX Arabidopsis thaliana.
OS
XX WO200142279-A2.
XX
XX 14-JUN-2001.
XX
XX 13-DEC-2000; 2000WO-GB004781.
XX
XX 13-DEC-1999; 99GB-00029469.
XX
XX (PROT-) PROTEOM LTD.
XX
XX Roberts GW, Heal JR;
XX
XX WPI; 2001-381629/40.
XX
XX A set of peptide ligands for agricultural research and development of
PT therapeutic agents comprise specific complementary peptides to proteins
PT encoded by genes of plant genomes.
XX
XX Example 4; Page 121; 201pp; English.
XX
XX The present invention relates to a set of peptide ligands consisting of
CC specific complementary peptides to proteins encoded by genes of plant
CC genomes. The present sequence is one such peptide from Arabidopsis
CC thaliana. The peptides of the present invention are useful in an assay to
CC identify a peptide, especially a peptide pesticide or herbicide. The
CC peptides are also useful for tools for agricultural research and
CC development
XX
XX Sequence 10 AA;
SQ

Query Match      1.3%; Score 8; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      208 GGGGSGSGA 215
Db      1 GGGGSGSGA 8

RESULT 31
AAG84110
ID AAG84110 standard; peptide; 10 AA.
XX
XX AAG84110;
AC
XX 11-SEP-2001 (first entry)
XX
XX Arabidopsis thaliana peptide ligand #750.
DE
XX Plant; peptide pesticide; peptide herbicide; agricultural research.
XX
XX Arabidopsis thaliana.
OS
XX WO200142279-A2.
XX
XX 14-JUN-2001.
XX
XX 13-DEC-2000; 2000WO-GB004781.
XX
XX 13-DEC-1999; 99GB-00029469.
XX
XX (PROT-) PROTEOM LTD.
XX
XX Roberts GW, Heal JR;
XX
XX WPI; 2001-381629/40.
XX
XX A set of peptide ligands for agricultural research and development of
PT therapeutic agents comprise specific complementary peptides to proteins
PT encoded by genes of plant genomes.
XX
XX Example 4; Page 135; 201pp; English.
XX
XX The present invention relates to a set of peptide ligands consisting of
CC specific complementary peptides to proteins encoded by genes of plant
CC genomes. The present sequence is one such peptide from Arabidopsis
CC thaliana. The peptides of the present invention are useful in an assay to
CC identify a peptide, especially a peptide pesticide or herbicide. The
CC peptides are also useful for tools for agricultural research and
CC development
XX
XX Sequence 10 AA;
SQ

Query Match      1.3%; Score 8; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      208 GGGGSGSGA 215
Db      2 GGGGSGSGA 9

RESULT 32
AAG84006
ID AAG84006 standard; peptide; 10 AA.
XX
XX AAG84006;
AC
XX 11-SEP-2001 (first entry)
XX
XX Arabidopsis thaliana peptide ligand #64.
DE
XX Plant; peptide pesticide; peptide herbicide; agricultural research.
XX
```

OS Arabidopsis thaliana.
XX
XX WO200142279-A2.
XX
XX 14-JUN-2001.
XX
XX 13-DEC-2000; 2000WO-GH004781.
XX
XX 13-DEC-1999; 99GB-00029469.
XX
XX (PROT-) PROTEOM LTD.
XX
XX Roberts GW, Heal JR;
XX
XX WPI; 2001-381629/40.
XX
XX A set of peptide ligands for agricultural research and development of
PT therapeutic agents comprise specific complementary peptides to proteins
PT encoded by genes of plant genomes.
XX
XX Example 4; Page 121; 201pp; English.
XX
XX The present invention relates to a set of peptide ligands consisting of
CC specific complementary peptides to proteins encoded by genes of plant
CC genomes. The present sequence is one such peptide from Arabidopsis
CC thaliana. The peptides of the present invention are useful in an assay to
CC identify a peptide, especially a peptide pesticide or herbicide. The
CC peptides are also useful for tools for agricultural research and
CC development
XX
XX Sequence 10 AA;
SQ

Query Match 1.3%; Score 8; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 208 GGGGGSGA 215
Db 2 GGGGGSGA 9

RESULT 33
AAR75542
ID AAR75542 standard; peptide; 15 AA.
XX
XX AAR75542;
AC
XX 25-MAR-2003 (revised)
DT 05-DEC-1995 (first entry)
XX
XX P. communis (pear) arabinogalactan protein AGP thermolysin peptide.
DE
XX Arabinogalactan protein; AGP; gum; thickening agent; emulsifier;
KW adhesive.
XX
XX Pyrus communis.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 1.15
FT /note= "P is hydroxyproline"
FT Misc-difference 15
FT /note= "residue is in parentheses"
XX
XX MO9515377-A1.
XX
XX 08-JUN-1995.
XX
XX 01-DEC-1994; 94WO-AU000744.
XX
XX 03-DEC-1993; 93US-00161944.
PR 18-JUN-1994; 94US-00276452.
XX
XX (ALBR) ALBRIGHT & WILSON AUSTRALIA.
PA

XX
XX Chen C, Mau S, Du H, Gane A, Bacic A, Clarke A;
PI
XX WPI; 1995-215258/28.
XX
XX DR
XX Recombinant plant arabinogalactan protein and cloned DNA encoding it -
PT useful as an emulsifying agent, adhesive agent or a lubricating agent.
XX
XX Claim 14; Page 65; 142pp; English.
XX
XX Total native arabinogalactan protein (AGP) was isolated from P. communis.
CC It was deglycosylated using anhydrous HF. The total deglycosylated AGP
CC sample was then reduced and carboxymethylated. It was then separated by
CC HPLC. The fractions R16.4, R18.2 and unbound were subjected to
CC thermolysin digestion and then separated. Individual peaks from
CC thermolysis digested R18.2 were sequenced. The sequences are given in
CC AAR75542-R75547 and are from peaks 1-5. AAR75547 is identical to the
CC sequence obtd. from peak 1 of R16.4. A cloned DNA molecule encoding a
CC protein backbone of an arabinogalactan protein from Pyrus which
CC hybridises to a nt. sequence encoding an AA sequence in AAR75542-R75547
CC is claimed (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 15 AA;
SQ

Query Match 1.3%; Score 8; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 66 SPAPSPAG 73
Db 4 SPAPSPAG 11

RESULT 34
ABP52300
ID ABP52300 standard; peptide; 15 AA.
XX
XX ABP52300;
AC
XX 16-OCT-2002 (first entry)
DT
XX
XX HLA-DR2 molecule binding peptide SEQ ID NO:94.
DE
XX
XX Human leukocyte antigen; HLA; major histocompatibility complex; MHC;
KW HLA-DR2; MHC class II DR-2 molecule; demyelinating; multiple sclerosis;
KW immune response; anti-inflammatory; neuroprotective; proliferation;
KW MHC class II protein inhibitor; demyelinating disease; inhibition;
KW post-viral encephalomyelitis; post-vaccine demyelinating condition;
KW anti-tumour necrosis factor agent.
XX
XX Homo sapiens.
OS
XX Synthetic.
OS
XX WO200259143-A2.
XX
XX 01-AUG-2002.
PD
XX 24-JAN-2002; 2002WO-US002071.
PF
XX 24-JAN-2001; 2001US-0263569P.
PR
XX
XX (HARD) HARVARD COLLEGE.
PA
XX Strominger JL, Fridkis-Hareli M;
PI
XX WPI; 2002-608439/65.
XX
XX New compositions comprising synthetic peptides in complex with a major
PT histocompatibility complex class II HLA-DR2 protein, useful for treating a
PT demyelinating disease, e.g. multiple sclerosis, or post-viral
PT encephalomyelitis.
XX
XX Claim 28; Page 39; 54pp; English.
PS

XX The present invention describes compositions (I) comprising a peptide
CC with an amino acid sequence with tyrosine (Y), lysine (K), and/or valine
CC residues. The complex of the peptide with a major histocompatibility
CC complex (MHC) class II HLA-DR2 protein is involved in modulating an
CC immune response. (I) has antiinflammatory and neuroprotective activities,
CC and can be used as a MHC class II protein inhibitor. The compositions
CC comprising the peptides are useful for treating demyelinating diseases
CC such as multiple sclerosis, post-viral encephalomyelitis, a post-vaccine
CC demyelinating condition, and a side effect of administering an anti-
CC tumour necrosis factor agents. The peptide further inhibits proliferation
CC of autotransgenic-specific HLA-DR2-restricted T cell clones. ABP52207 to
CC ABP52305 represent peptides used in the exemplification of the present
CC invention
XX
SQ Sequence 15 AA;

Query Match 1.3%; Score 8; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 8 FRAAAAPA 15
Db 8 FRAAAAPA 15

RESULT 35
ADFA1646
ID ADFA1646 standard; peptide; 15 AA.
AC ADFA1646;
XX
XX
XX

DT 12-FEB-2004 (first entry)

DE Human macroprotein-49.61 N-terminal peptide.

XX Human; macroprotein-49.61; recombinant production; gene therapy;
KM cleft lip; cleft palate; N-terminal peptide;
KM enzyme linked immunosorbent assay; ELISA.
XX
XX

OS Homo sapiens.

PN CN1381500-A.

PD 27-NOV-2002.

PF 18-APR-2001; 2001CN-00112653.

PR 18-APR-2001; 2001CN-00112653.

XX (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.

PI Mao Y, Xie Y;

XX WPI; 2003-258244/26.

PT Polypeptide-human macroprotein -49.61 and polynucleotide for coding it.

XX Example 6; SEQ ID NO 7; 31pp; Chinese.

CC The invention relates to human macroprotein-49.61 (ADFA1641) and nucleic
CC acids encoding it (ADFA1640). The protein has a molecular weight of 49.61
CC kD. The invention also relates to a method for the recombinant production
CC of the protein, an antagonist of the protein, and the use of the protein,
CC gene and antagonist in therapeutic applications. Macroprotein-49.61 can
CC be used in the treatment of a variety of conditions such as cleft lip and
CC cleft palate. The present sequence represents the 15 N-terminal amino
CC acids of human macroprotein-49.61 used in ELISA (enzyme linked
CC immunosorbent assay) in an exemplification of the invention.
XX
SQ Sequence 15 AA;

Query Match 1.3%; Score 8; DB 7; Length 15;

Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 FSEIEDL 423
Db 8 FSEIEDL 15

RESULT 36
ABP52302
ID ABP52302 standard; peptide; 17 AA.
XX
XX

AC ABP52302;

DT 16-OCT-2002 (first entry)

DE HLA-DR2 molecule binding peptide SEQ ID NO:96.

XX Human leukocyte antigen; HLA; major histocompatibility complex; MHC;
KM HLA-DR2; MHC class II DR-2 molecule; demyelinating; multiple sclerosis;
KM immune response; antiinflammatory; neuroprotective; proliferation;
KM MHC class II protein inhibitor; demyelinating disease; inhibition;
KM post-viral encephalomyelitis; post-vaccine demyelinating condition;
XX anti-tumour necrosis factor agent.
XX

OS Homo sapiens.
OS Synthetic.

PN WO200259143-A2.

PD 01-AUG-2002.

PF 24-JAN-2002; 2002WO-US002071.

PR 24-JAN-2001; 2001US-0263569P.

XX (HARD) HARVARD COLLEGE.

PI Strominger JL, Fridkis-Hareli M;

XX WPI; 2002-608439/65.

XX New compositions comprising synthetic peptides in complex with a major
PT histocompatibility complex class II HLA-DR2 protein, useful for treating a
PT demyelinating disease, e.g. multiple sclerosis, or post-viral
XX encephalomyelitis.
XX

PS Claim 28; Page 39; 54pp; English.

CC The present invention describes compositions (I) comprising a peptide
CC with an amino acid sequence with tyrosine (Y), lysine (K), and/or valine
CC residues. The complex of the peptide with a major histocompatibility
CC complex (MHC) class II HLA-DR2 protein is involved in modulating an
CC immune response. (I) has antiinflammatory and neuroprotective activities,
CC and can be used as a MHC class II protein inhibitor. The compositions
CC comprising the peptides are useful for treating demyelinating diseases
CC such as multiple sclerosis, post-viral encephalomyelitis, a post-vaccine
CC demyelinating condition, and a side effect of administering an anti-
CC tumour necrosis factor agents. The peptide further inhibits proliferation
CC of autotransgenic-specific HLA-DR2-restricted T cell clones. ABP52207 to
CC ABP52305 represent peptides used in the exemplification of the present
CC invention
XX
SQ Sequence 17 AA;

Query Match 1.3%; Score 8; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 FRAAAAPA 15
Db 10 FRAAAAPA 17

CC pathological status, or pathology sources such as tumours, or diagnosing
CC or characterising mitochondrial function in patients with diseases,
CC especially genetic or metabolic diseases (all claimed). The method is
CC useful for studying the regulation of mitochondrial function in vitro, to
CC decipher cell death mechanisms and to identify new therapeutic drugs. The
CC peptides are capable of inducing apoptosis in cultured cells including
CC adenocarcinoma cell lines and human umbilical vein embryonic cells and
CC therefore have potential in treating human diseases. Mitochondrio-toxic
CC or mitochondrio-protective drugs identified by the method are useful for
CC treating cancers and autoimmune diseases, acute or chronic ischaemia,
CC obesity and diabetes.
CC
CC
SQ Sequence 31 AA;

Query Match 1.3%; Score 8; DB 8; Length 31;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
Db 10 GGGGGSGA 17

RESULT 39
ADN11207
ID ADN11207 standard; peptide; 31 AA.
AC ADN11207;
DT 01-JUL-2004 (first entry)
DE Peptide modulator of mitochondrial function.
XX Mitochondria; cytostatic; vasotropic; antidiabetic.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 6 /note= "Advantageously replaced by D-Tyr"
XX
XX W02004031768-A2.
XX
XX 15-APR-2004.
XX
XX 02-OCT-2003; 2003WO-BP012056.
XX
XX 02-OCT-2002; 2002US-0415092P.
XX 23-MAY-2003; 2003US-0472725P.
XX
XX (THER-) THERAPOTOSIS SA.
XX
XX Jacotot E, Lecoeur H, Rebouillat D;
XX
XX WPI; 2004-348022/32.
XX
XX
XX Screening modulators of mitochondrial function, for treating e.g.,
XX cancer, comprises adding a compound to a mitochondria preparation and
XX using fluorimetric analysis of morphology, morphometric parameters, and
XX membrane integrity.
XX
XX
XX Claim 12; SEQ ID NO 28; 45pp; English.
XX
XX The present sequence is that of a peptide modulator of mitochondrial
XX function. It is an example of peptides that contain 8-50 amino acids and
XX have the following features: at least a part of the peptide is an
XX amphipathic alpha helix; 2-4 amino acids are positively charged (lys or
XX arg) and are located on the same side of the helix; when the helix is
XX projected using helical wheel representation, the Arg and/or Lys residues
XX form a cluster; and when added to purified, isolated mitochondria they
XX induce changes in ultrastructure or membrane permeability. Such peptides
XX can be identified using the method of the invention. This involves adding
XX a test compound to a purified, isolated mitochondria preparation and

CC simultaneously using a real-time fluorimetric analysis of mitochondrial
CC morphology combining analysis of morphometric parameters (side scatter
CC (SSC)/forward scatter (FSC) parameters) with analysis of membrane
CC integrity by dye fluorescence. Claimed peptides ADN1180-ADN11214
CC preferably comprise a stabilising group such as an amide, alkyl or acyl
CC group in the C- and N-terminal positions, and a marker or linker group.
CC They can include L- or D-amino acids retro-inverso, have a reduced
CC peptidic backbone or be pseudo peptides. The peptides are bound to a
CC peptide delivery system such as a targeting sequence, an antibody or its
CC fragment (scfv). A linker between the peptide and delivery system may
CC allow the peptide to adopt a helical structure. The method is useful for
CC designing and defining new agents aimed at modulating mitochondrial
CC membrane permeability and defining and manufacturing reagents or kits of
CC reagent to put the method in use. It is also useful for characterising
CC mitochondrial reactivity in specific mitochondria, to be used to
CC ascertain mitochondrial function in various tissues of physiological or
CC pathological status, or pathology sources such as tumours, or diagnosing
CC or characterising mitochondrial function in patients with diseases,
CC especially genetic or metabolic diseases (all claimed). The method is
CC useful for studying the regulation of mitochondrial function in vitro, to
CC decipher cell death mechanisms and to identify new therapeutic drugs. The
CC peptides are capable of inducing apoptosis in cultured cells including
CC adenocarcinoma cell lines and human umbilical vein embryonic cells and
CC therefore have potential in treating human diseases. Mitochondrio-toxic
CC or mitochondrio-protective drugs identified by the method are useful for
CC treating cancers and autoimmune diseases, acute or chronic ischaemia,
CC obesity and diabetes.
CC
CC
SQ Sequence 31 AA;

Query Match 1.3%; Score 8; DB 8; Length 31;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
Db 10 GGGGGSGA 17

RESULT 40
ADN11226
ID ADN11226 standard; peptide; 31 AA.
AC ADN11226;
DT 01-JUL-2004 (first entry)
DE Peptide modulator of mitochondrial function.
XX Mitochondria; cytostatic; vasotropic; antidiabetic.
XX Synthetic.
XX
XX W02004031768-A2.
XX
XX 15-APR-2004.
XX
XX 02-OCT-2003; 2003WO-BP012056.
XX
XX 02-OCT-2002; 2002US-0415092P.
XX 23-MAY-2003; 2003US-0472725P.
XX
XX (THER-) THERAPOTOSIS SA.
XX
XX Jacotot E, Lecoeur H, Rebouillat D;
XX
XX WPI; 2004-348022/32.
XX
XX
XX Screening modulators of mitochondrial function, for treating e.g.,
XX cancer, comprises adding a compound to a mitochondria preparation and
XX using fluorimetric analysis of morphology, morphometric parameters, and
XX membrane integrity.
XX

PS Disclosure, SEQ ID NO 47, 45bp, English.

XX The present sequence is that of a peptide modulator of mitochondrial
CC function. The peptide is capable of inducing cell death of adenocarcinoma
CC cell lines. Such peptides are identified using the method of the
CC invention. This involves adding a test compound to a purified, isolated
CC mitochondrial preparation and simultaneously using a real-time
CC fluorimetric analysis of mitochondrial morphology combining analysis of
CC morphometric parameters (side scatter (SSC)/forward scatter (FSC)
CC parameters) with analysis of membrane integrity by dye fluorescence.
CC Preferred peptides ADN1180-ADN11214 may have a stabilising group, such as
CC an amide, alkyl or acyl group in the C- and N-terminal positions, and a
CC marker or linker group. They can include L- or D-amino acids retro-
CC inverso, have a reduced peptidic backbone or be pseudo peptides. The
CC peptides are bound to a peptidic delivery system such as a targeting
CC sequence, an antibody or its fragment (scFv). A linker between the
CC peptide and delivery system may allow the peptide to adopt a helical
CC structure. The method is useful for designing and defining new agents
CC aimed at modulating MMP and defining and manufacturing reagents or kits
CC of reagent to put the method in use. It is also useful for characterising
CC mitochondrial reactivity in specific mitochondria, to be used to
CC ascertain mitochondrial function in various tissues of physiological or
CC pathological status, or pathology sources such as tumours, or diagnosing
CC or characterising mitochondrial function in patients with diseases,
CC especially genetic or metabolic diseases (all claimed). The method is
CC useful for studying the regulation of mitochondrial function in vitro, to
CC decipher cell death mechanisms and to identify new therapeutic drugs. The
CC peptides are capable of inducing apoptosis in cultured cells including
CC adenocarcinoma cell lines and human umbilical vein embryonic cells and
CC therefore have potential in treating human diseases. Mitochondrio-toxic
CC or mitochondrio-protective drugs identified by the method are useful for
CC treating cancers and autoimmune diseases, acute or chronic ischaemia,
CC obesity and diabetes.

SO Sequence 31 AA:

Query Match 1.3%; Score 8; DB 8; Length 31;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215

DB 10 GGGGGSGA 17

RESULT 41

AA02201 standard; protein; 64 AA.

AC AAG02201;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 6282.

KM Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0020610.

PR 26-FEB-1999; 99US-0122487P.

PA (GENST) GENSET.

PI Dumas Milne Edwards J, Duclet A, Giordano J;

DR WPI; 2000-500381/45.

DR N-PSDB; AAC02207.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 13; SEQ ID NO 6282; 71bp + Sequence Listing; English.

XX The present sequence is a polypeptide encoded by one of a large number of
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC tissues. EST sequences usually correspond mainly to the 3' untranslated
CC region (UTR) of the mRNA because they are often obtained from oligo-dT
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC sequences derived from the 5' ends of mRNAs and even in those cases where
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC are also used in diagnostic, forensic, gene therapy and chromosome
CC mapping procedures. They are used to obtain upstream regulatory sequences
CC and to design expression and secretion vectors

SO Sequence 64 AA:

Query Match 1.3%; Score 8; DB 3; Length 64;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 416 FSSRIEDL 423

DB 24 FSSRIEDL 31

RESULT 42

ABP10891 standard; protein; 79 AA.

AC ABP10891;

DT 24-JUN-2002 (first entry)

DE Human ORFX protein sequence SEQ ID NO:21764.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KM hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KM degenerative disorder; osteoarthritis; neurodegenerative disorder;
KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KM hypertension; hypothyroidism; cholesterol ester storage disease;
KM immune deficiency; immune disorder; infectious disease;
KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KM myasthenia gravis.

OS Homo sapiens.

PN WO200192523-A2.

PD 06-DEC-2001.

PF 29-MAY-2001; 2001WO-US010836.

PR 30-MAY-2000; 2000US-0206132P.

PR 29-AUG-2000; 2000US-0228716P.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach MD;

DR WPI; 2002-106308/14.

DR N-PSDB; ABN26643.

PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.

XX WPI; 1997-289413/26.
 DR
 XX
 PT Diagnosing Trypanosoma cruzi infection by detecting antibodies to novel
 PT antigens - which are useful in vaccines to provide protective immunity
 PT against Chagas' disease.
 XX
 XX Example 7; Page 87; 110pp; English.
 CC This synthetic polypeptide sequence, designated 2/Lo/2E/D, comprises a
 CC multi-epitope combination of the Trypanosoma cruzi PRP2 epitope (see
 CC AAM19066), TcLo1.2 epitope (see AAM19083), TcE epitope (see AAT19062) and
 CC TcD epitope (see AAM19064) in linear sequence. It was produced by
 CC combining synthetic PRP2/TcLo1.2 and TcD/TcE dipeptides. A claimed method
 CC for detecting T. cruzi infection in a biological sample involves
 CC contacting the sample with such a combination polypeptide, and detecting
 CC the presence of antibodies that bind to the polypeptide
 XX
 SQ Sequence 83 AA;
 Query Match 1.3%; Score 8; DB 2; Length 83;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 8; Conservative 0; Mismatches 0; Gaps 0;
 Indels 0;
 QY 9 KAAAPAG 16
 |||||
 56 KAAAPAG 63
 Db
 RESULT 45
 AAY23318
 ID AAY23318 standard; peptide; 83 AA.
 XX
 XX AAY23318;
 AC
 XX 31-AUG-1999 (first entry)
 DT
 XX
 DE Peptide containing four immunoreactive T. cruzi epitopes.
 XX
 KM Trypanosoma cruzi epitope; Trypanosoma cruzi infection; antigen; vaccine;
 KM Chagas' disease.
 XX
 OS Synthetic.
 OS Trypanosoma cruzi.
 XX
 PN W09931246-A1.
 XX
 PD 24-JUN-1999.
 XX
 PF 04-DEC-1998; 98WO-US025871.
 XX
 PR 18-DEC-1997; 97US-00993674.
 XX
 PA (CORI-) CORIXA CORP.
 PI Reed SG, Skeiky YAM, Lodes MJ, Houghton RL, Smith JM, McNeill PD;
 XX WPI; 1999-405035/34.
 DR
 XX
 PT New isolated Trypanosoma cruzi epitopes.
 XX
 PS Example 7; Page 91; 103pp; English.
 CC The specification describes new Trypanosoma cruzi epitopes. A method for
 CC detecting Trypanosoma cruzi infection in a biological sample comprises
 CC contacting the sample with a polypeptide comprising an epitope of a TC
 CC antigen, or a variant of the antigen that differs only in conservative
 CC substitutions and/or modifications and detecting the presence of
 CC antibodies that bind to the polypeptide in the sample, thereby detecting
 CC infection. The TC polypeptides can be used in vaccines for inducing
 CC protective immunity against Chagas' disease in a patient. The
 CC polypeptides and antibodies can also be used for detecting TC infection.
 CC The present sequence represents a peptide containing four immunoreactive

CC T. cruzi epitopes from PRP-2, TcD, TcE and TcLo1.2 in a linear sequence
 XX
 SQ Sequence 83 AA;
 Query Match 1.3%; Score 8; DB 2; Length 83;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 8; Conservative 0; Mismatches 0; Gaps 0;
 Indels 0;
 QY 9 KAAAPAG 16
 |||||
 56 KAAAPAG 63
 Db
 RESULT 46
 AAM6330
 ID AAM6330 standard; protein; 98 AA.
 XX
 AC AAM6330;
 AC
 XX 01-MAR-1999 (first entry)
 DT
 XX
 DE Kidney injury associated molecule HM074 protein.
 XX
 KM Kidney injury associated molecule; kidney injury related molecule; KIM;
 KM tissue growth promotion; regeneration; renal condition;
 KM acute renal failure; acute nephritis; tumour.
 XX
 OS Rattus sp.
 XX
 PN W09853071-A1.
 XX
 PD 26-NOV-1998.
 XX
 PF 22-MAY-1998; 98WO-US010547.
 XX
 PR 23-MAY-1997; 97US-0047490P.
 PR 23-MAY-1997; 97US-0047491P.
 PA (BIOJ) BIOGEN INC.
 XX
 XX Sanicola-Nadel M, Hession CA, Wei H, Cate RL;
 PI
 DR WPI; 1999-045312/04.
 DR N-PSDB; AAV80618.
 XX
 XX Kidney injury-associated molecule, KIM, polypeptides - upregulated in
 PT injured or regenerating tissues, useful to promote tissue growth and
 PT regeneration, especially to treat renal conditions.
 XX
 PS Claim 17; Page 153; 213pp; English.
 CC The present sequence represents a kidney injury associated molecule (KIM)
 CC protein. KIM proteins can be administered therapeutically by expressing
 CC KIM encoding polynucleotides, to promote growth and/or survival of
 CC damaged tissue (e.g. renal tissue), since the KIM proteins are
 CC upregulated in injured or regenerating (especially renal) tissues. KIM
 CC fusion proteins, conjugates, antibodies and vectors can also be used
 CC therapeutically, e.g. these or the KIM proteins may be included with an
 CC acceptable carrier in pharmaceutical compositions, useful for therapy/
 CC prophylaxis of conditions associated with dysfunction/dysregulation of
 CC KIM genes or proteins, especially renal diseases or impairments of renal
 CC function in humans (e.g. acute renal failure, acute nephritis). The
 CC polynucleotides can be used to produce antisense sequences which, when
 CC internalised into cells, can disrupt expression of a cellular KIM gene,
 CC also useful in therapy (e.g. to block the growth of tumours dependent on
 CC KIM for growth) or compositions. The proteins and polynucleotides are
 CC useful diagnostically e.g. to detect and quantify renal injury/disease
 CC (indicative of increased risk, or presence of, renal injury or impaired
 CC function), or abnormal responses to tissue injury (indicative of
 CC increased risk, or presence of, an autoimmune response or abnormal tissue
 CC growth arising from/affecting renal tissue). The proteins can also be
 CC used to locate KIM-producing cells (especially specific loci, e.g. tissue
 CC masses abnormally producing/expressing KIM such as tumours arising

CC from/affecting renal tissue), by contacting cells with an imageable KIM-
CC binding reagent and imaging reagent accumulation
XX
SQ Sequence 98 AA;
Query Match 1.3%; Score 8; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 386 TPEKPRSS 393
Db 42 TPEKPRSS 49
RESULT 47
AAG34350
ID AAG34350 standard; protein; 109 AA.
XX
AC AAG34350;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 41779.
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147719P.
PR 05-AUG-1999; 99US-0147260P.

PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149175P.
PR 17-AUG-1999; 99US-0149426P.
PR 18-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158222P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.3%; Score 8; DB 3; Length 109;
Best Local Similarity 100.0%; Pred. No. 91;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 AAAAPAG 17
Db 72 AAAAPAG 79

RESULT 48
AAE35343
ID AAE35343 standard; protein; 109 AA.
XX
AC AAE35343;
XX
DT 17-JUN-2003 (first entry)
XX
DE Human colon cancer-associated polypeptide, CO-94 (seb4D).
XX
KW Human; colon cancer-associated polypeptide; immune response; therapy;
XX colon cancer.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 84 /label= Unknown
FT /note= "Encoded by NGA"
FT Misc-difference 100 /label= Unknown
FT /note= "Encoded by ANA"
XX
XX W0200290986-A1.
XX PD 14-NOV-2002.
XX PF 02-MAY-2002; 2002MO-US013994.
XX PR 04-MAY-2001; 2001US-00849602.
XX
XX (LUDWIG-) LUDWIG INST CANCER RES.
XX (CORR) CORNELL RES FOUND INC.
XX
XX Chen Y, Old LJ, Scanlan MJ, Stockert E;
XX WPI; 2003-112003/10.
XX DR N-PSDB; AAD54022.
XX
XX Diagnosing colon cancer in a subject comprises identifying colon cancer-
XX associated polypeptides as antigens that elicit immune response in colon
XX cancer.
XX
XX Example 1; Page 101; 122pp; English.
XX
XX The invention relates to a method for diagnosing colon cancer in a
XX subject which comprises identifying colon cancer-associated polypeptides
XX as antigens that elicit immune response in colon cancer. The method is
XX useful for diagnosing, determining onset, progression, or regression of
XX colon cancer in a subject, or for selecting a course of treatment of a
XX subject having or suspected of having colon cancer. The colon cancer-
XX associated polypeptides are useful as markers for diagnosing colon
XX cancer, and for following the course of treatment of colon cancer. The
XX present sequence is human colon cancer-associated polypeptide
XX Sequence 109 AA;
SQ
Query Match 1.3%; Score 8; DB 6; Length 109;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 160 ASPAATAS 167
Db 102 ASPAATAS 109

RESULT 49
AAM43554
ID AAM43554 standard; protein; 112 AA.
XX
AC AAM43554;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 232.
XX
XX Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebrotrophic; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cyostatic; immunosuppressive; nootropic;
KW neuroprotective; anti allergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; vulnerable; anticonvulsant; antibacterial;
KW antiparasitic; cardiac; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human.
OS Homo sapiens.
XX
XX MO200155308-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US001309.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-020515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0235484P.
PR 25-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240360P.
PR 20-OCT-2000; 2000US-0241211P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.

PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259567B.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-488781/53.
XX N-PSDB; AA163860.
DR
XX New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders.
XX
PS Claim 11; SEQ ID NO 232; 664pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AA163803-AA164012) and
CC the encoded proteins (AA434497-AA43660) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 112 AA;
Query Match 1.3%; Score 8; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 321 SCGRSGKG 328
Db 41 SCGRSGKG 48
RESULT 50
ADM24575
ID ADM24575 standard; protein; 112 AA.
XX
XX ADM24575;
AC
XX 20-MAY-2004 (first entry)
XX
XX Human PRO polypeptide #58.
DE
XX Human PRO polypeptide; transmembrane polypeptide;
KM Human; PRO; secreted polypeptide; immune system disorder;
KM neuro disorder; gastrointestinal disorder; muscular disorder;
KM reproductive disorder; gastrointestinal disorder; pulmonary disorder;
KM cardiovascular disorder; renal disorder; proliferative disorder; cancers;
KM systemic lupus erythematosus; rheumatoid arthritis; multiple sclerosis;
PR

KM thyroiditis; anaemia; Grave's disease; diabetes; hepatitis; asthma;
KM allergy; nephritis; Parkinson's disease; Alzheimer's disease;
KM atherosclerosis; myocardial infarction; AIDS; infection.
XX
XX Homo sapiens.
XX US2004014039-A1.
PN
PD 22-JAN-2004.
XX
XX 31-MAY-2002; 2002US-00158057.
PF
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 26-SEP-2000; 2000US-0234988P.
 PR 27-SEP-2000; 2000US-0235844P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246533P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249247P.
 PR 17-NOV-2000; 2000US-0249248P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 01-DEC-2000; 2000US-0250393P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PR 17-JAN-2001; 2001US-00764890.
 XX
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Ruben SM, Barash SC;
 XX WPI, 2004-108205/11.
 DR N-PSDB; ADM24411.
 XX
 XX
 PT Secreted and transmembrane PRO polypeptides and nucleic acids for
 PT diagnosing, preventing or treating diseases associated with aberrant
 PT expression or activity of the polypeptide, e.g. cancer, asthma, AIDS,
 PT Parkinson's disease or diabetes.
 XX
 XX
 PS Claim 11; SEQ ID NO 232; 284bp; English.
 XX
 CC The invention relates to a secreted and transmembrane PRO polypeptides
 CC and the polynucleotides encoding them. The polypeptides and
 CC polynucleotides are useful in diagnosing, preventing, prognosing or
 CC treating diseases or disorders associated with aberrant expression and/or
 CC activity of PRO polypeptides, such as neural disorders, immune system
 CC disorders, muscular disorders, reproductive disorders, gastrointestinal
 CC disorders, pulmonary disorders, cardiovascular disorders, renal
 CC disorders, proliferative disorders and/or cancers. In particular, these
 CC diseases are systemic lupus erythematosus, rheumatoid arthritis, multiple
 CC sclerosis, thyroiditis, anaemia, Grave's disease, diabetes, hepatitis,
 CC asthma, allergies, nephritis, Parkinson's disease, Alzheimer's disease,
 CC atherosclerosis, myocardial infarction, AIDS, infections, etc. This
 CC sequence represents a human PRO polypeptide of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 112 AA:
 QY 321 SCGRSGKG 328
 Db 41 SCGRSGKG 48
 Query Match 1.3%; Score 8; DB 8; Length 112;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 51
 ID AAG41132 standard; protein; 115 AA.
 XX
 AC AAG41132;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Zea mays protein fragment SEQ ID NO: 51135.
 XX
 XX Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 KM termination sequence; corn.
 XX
 OS Zea mays subsp. mays.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-00301439.
 XX
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134370P.
PR 14-MAY-1999; 99US-0134768P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136382P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139839P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142058P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.

PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 27-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153707P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 05-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.

```
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161932P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.
```

```
Query Match 1.3%; Score 8; DB 3; Length 115;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 10 AAAAPAGG 17
    |||||
Db 72 AAAAPAGG 79
```

```
RESULT 52
AAG01944
ID AAG01944 standard; protein; 123 AA.
```

```
XX AAG01944;
```

```
DT 06-OCT-2000 (first entry)
```

```
DE Human secreted protein, SEQ ID NO: 6025.
```

```
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
```

```
XX Homo sapiens.
```

```
OS Homo sapiens.
```

```
PN EP1033401-A2.
```

```
PD 06-SEP-2000.
```

```
PF 21-FEB-2000; 2000BP-00200610.
```

```
PR 26-FEB-1999; 99US-0122487P.
```

```
XX (GEST ) GENSET.
```

```
PI Dumas Milne Edwards J, Duclert A, Giordano J;
```

```
DR WPI, 2000-500381/45.
```

```
XX N-PSDB; AAC01950.
```

```
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
```

```
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
```

```
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
```

```
PS Claim 13; SEQ ID NO 6025; 71bp + Sequence Listing; English.
XX The present sequence is a polypeptide encoded by one of a large number of
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC tissues. EST sequences usually correspond mainly to the 3' untranslated
CC region (UTR) of the mRNA because they are often obtained from oligo-dT
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC sequences derived from the 5' ends of mRNAs and even in those cases where
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC are also used in diagnostic, forensic, gene therapy and chromosome
CC mapping procedures. They are used to obtain upstream regulatory sequences
CC and to design expression and secretion vectors
```

```
OY 416 FSESEIDL 423
    |||||
Db 51 FSESEIDL 58
```

```
RESULT 53
ABG19967
ID ABG19967 standard; protein; 140 AA.
```

```
XX ABG19967;
```

```
DT 13-FEB-2002 (first entry)
```

```
DE Novel human diagnostic protein #19958.
```

```
XX Human; chromosome mapping; gene mapping; forensic;
```

```
XX food supplement; medical imaging; diagnostic; genetic disorder.
```

```
XX Homo sapiens.
```

```
PN WO200175067-A2.
```

```
PD 11-OCT-2001.
```

```
PF 30-MAR-2001; 2001MO-US008631.
```

```
PR 31-MAR-2000; 2000US-00540217.
```

```
PR 23-AUG-2000; 2000US-00649167.
```

```
XX (HYSE-) HYSEQ INC.
```

```
PI Drmanac RT, Liu C, Tang YT;
```

```
DR WPI, 2001-639362/73.
```

```
XX N-PSDB; AAS84154.
```

```
XX New isolated polynucleotide and encoded polypeptides, useful in
```

```
PT diagnostics, forensics, gene mapping, identification of mutations
```

```
PT responsible for genetic disorders or other traits and to assess
```

```
PT biodiversity.
```

```
XX Claim 20; SEQ ID NO 50326; 103bp; English.
```

```
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
```

```
CC sequences. (I) is useful as hybridisation probes, polymerase chain
```

```
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
```

```
CC and in recombinant production of (II). The polynucleotides are also used
```

```
CC in diagnostics as expressed sequence tags for identifying expressed
```

```
CC genes. (I) is useful in gene therapy techniques to restore normal
```

```
CC activity of (II) or to treat disease states involving (II). (II) is
```

```
CC useful for generating antibodies against it, detecting or quantitating a
```

CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 140 AA;

Query Match 1.3%; Score 8; DB 4; length 140;
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Gaps 0;
Matches 8; Conservative 0; Indels 0;

QY 239 QLAKEKE 246
Db 130 QLAKEKE 137

RESULT 54
AAV93408
ID AAV93408 standard; protein; 142 AA.
XX
AC AAV93408;
XX
DT 04-SEP-2000 (first entry)
XX
DE Amino acid sequence of a MPPI binding factor-1 polypeptide.
XX
XX MPPI binding factor 1; MAP1; transgenic plant; chimeric gene; NMPI;
XX nuclear matrix protein-1; FLIP; filament-like protein;
XX transcriptional activator; protein composition; crop growth;
XX crop protection.
XX
XX Zea mays.
XX
XX WO200028054-A2.
XX
PD 18-MAY-2000.
XX
PF 04-NOV-1999; 99WO-US025993.
XX
XX 06-NOV-1998; 98US-00187999.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E. I.
XX
PI Gindullis F, Meier I;
XX
DR MPPI; 2000-376565/32.
XX
XX N-PSDB; AAA46485.
XX
PT Regulating gene expression in plant cells, useful e.g. for expressing new
PT traits, by introducing genes for transcriptional activators that interact
PT with nuclear matrix proteins.
XX
XX
PS Claim 20; Page 75; 82pp; English.
XX
CC The present sequence represents a MPPI binding factor 1, designated MAP1.
CC It is used in the course of the invention. The specification describes a
CC method whereby gene expression in a stably transformed transgenic plant
CC cell is regulated using a chimeric gene that encodes NMPI (nuclear matrix
CC protein-1) or FLIP (filament-like protein) 4. Two chimeric genes are
CC combined into the genome, the first gene comprising a promoter linked to
CC a DNA binding domain, a coding sequence or its complement, and a DNA
CC polyadenylation sequence. The second gene comprises a promoter, a DNA
CC binding domain, NMPI or FLIP polynucleotides, and a polyadenylation
CC sequence. Expression of the first gene regulates expression of the

CC second, NMPI, FLIP4 and related proteins are parts of the nuclear
CC skeleton and function as transcriptional activators. The methods are used
CC to regulate gene expression in plants, e.g. to express novel traits in
CC transgenic plants that may result in new products useful in foods,
CC pharmaceuticals and materials, or for suppressing endogenous genes to
CC alter the protein composition, and to derive new phenotypes beneficial
CC for crop growth and development. NMPI, FLIP4 and related proteins are
CC used to alter level of expression of MPPI- or MAP1-binding proteins in
CC plants by co-suppression or overexpression, or to screen for compounds
CC that inhibit activity of MPPI- or MAP1-binding proteins, which are
CC potentially useful as crop protection agents. They can also be used to
CC study the plant nucleus matrix
XX
SQ Sequence 142 AA;

Query Match 1.3%; Score 8; DB 3; length 142;
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Gaps 0;
Matches 8; Conservative 0; Indels 0;

QY 10 AAAAPAGG 17
Db 13 AAAAPAGG 20

RESULT 55
AAG41131
ID AAG41131 standard; protein; 143 AA.
XX
AC AAG41131;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 51134.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence; corn.
XX
XX Zea mays subsp. mays.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX
XX 05-MAR-1999; 99US-0123180P.
XX
XX 09-MAR-1999; 99US-0123548P.
XX
XX 23-MAR-1999; 99US-0125788P.
XX
XX 25-MAR-1999; 99US-0126264P.
XX
XX 29-MAR-1999; 99US-0126785P.
XX
XX 01-APR-1999; 99US-0127462P.
XX
XX 06-APR-1999; 99US-0128234P.
XX
XX 08-APR-1999; 99US-0128714P.
XX
XX 16-APR-1999; 99US-0129845P.
XX
XX 19-APR-1999; 99US-0130077P.
XX
XX 21-APR-1999; 99US-0130449P.
XX
XX 23-APR-1999; 99US-0130510P.
XX
XX 28-APR-1999; 99US-0130891P.
XX
XX 30-APR-1999; 99US-0131449P.
XX
XX 30-APR-1999; 99US-0132048P.
XX
XX 04-MAY-1999; 99US-0132484P.
XX
XX 05-MAY-1999; 99US-0132485P.
XX
XX 06-MAY-1999; 99US-0132486P.
XX
XX 07-MAY-1999; 99US-0132487P.
XX
XX 11-MAY-1999; 99US-0132863P.
XX
XX 14-MAY-1999; 99US-0134256P.
XX
XX 14-MAY-1999; 99US-0134218P.
XX
XX 14-MAY-1999; 99US-0134219P.
XX
XX 14-MAY-1999; 99US-0134221P.
XX
XX 14-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.

PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147316P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150684P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154479P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0156599P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157665P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.

PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.3%; Score 8; DB 3; Length 143;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAAPAGG 17
Db 100 AAAAPAGG 107

RESULT 56
AAB92594
ID AAB92594 standard; protein; 146 AA.

AC AAB92594;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:10837.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

XX Claim 8; SEQ ID NO 10837; 2537bp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:

CC (a) an oligo-dT primer and an oligonucleotide complementary to the

CC complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in the

CC specification. The primer sets can be used in antisense therapy and in

CC gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH15629 to AAH1632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

QY Sequence 146 AA;
Best Local Similarity 1.3%; Score 8; DB 4; Length 146;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 EQQQQQQLQ 241
Db 112 EQQQQQQLQ 119

RESULT 57
ABP43613
ID ABP43613 standard; protein; 146 AA.

AC ABP43613;

DT 26-FEB-2003 (first entry)

DE FLJ10193 file clone HEMBA1004763.

KW Neuroprotective; immunomodulator; cancer; chromosome 17; cytostatic;

KW anti-inflammatory; gene therapy; nutritional supplement; wound; burn;

KW ulcer; Alzheimer's disease; Huntington's disease;

KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;

KW vulnerable.

OS Homo sapiens.

PN WO200231111-A2.

PD 18-APR-2002.

PF 11-OCT-2001; 2001WO-US027760.

PF 12-OCT-2000; 2000US-00687527.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Aundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Dymnac RT;

DR WPI; 2002-426278/45.

DR N-PSDB; ABQ60857.

XX Claim 20; SEQ ID # 516; 357bp + Sequence Listing; English.

CC The invention relates to 446 newly isolated polynucleotide sequences. The

CC activity of polynucleotides of the invention may be described as,

CC inflammatory, neuroprotective, immunomodulator, cytostatic and anti-

CC useful for treating a mammalian subject, or as nutritional sources or

CC supplements. These are useful in gene therapy, particularly for treating

CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,

CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or

CC inflammation. The nucleic acids and polypeptides are also useful in

CC diagnostic and research methods. The sequences given in record ABP43544-

CC ABP4389 represent polypeptides encoded by polynucleotides of the

CC invention. NOTE: The sequence data for this patent did not form part of

CC the printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX	Sequence	146 AA;
QY	Query Match	1.3%; Score 8; DB 5; Length 146;
	Best Local Similarity	100.0%; Pred. No. 1.2e+02;
Matches	8; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	234 E00000LOL 241	
	112 E00000LOL 119	
RESULT 58		
ID	ABM81840	standard; protein; 146 AA.
XX	ABM81840;	
DT	18-NOV-2004	(first entry)
DE	Tumour-associated antigenic target (TAT) polypeptide PRO82695,	SEQ:4738.
XX	Tumour-associated antigenic target; TAT; human; overexpression; cancer;	
KW	tumour; diagnosis; cell proliferative disorder; breast cancer;	
KW	colorectal cancer; lung cancer; ovarian cancer; liver cancer;	
KW	central nervous system cancer; bladder cancer; pancreatic cancer;	
KW	cervical cancer; melanoma; leukaemia; hybridisation probe;	
KW	chromosome identification; chromosome mapping; gene mapping;	
KW	gene therapy; cytostatic.	
OS	Homo sapiens.	
PN	WO2004030615-A2.	
PD	15-APR-2004.	
PF	29-SEP-2003; 2003WO-US028547.	
PR	02-OCT-2002; 2002US-0414971P.	
PA	(GETH) GENENTECH INC.	
XX	Wu TD, Zhang Z, Zhou Y;	
DR	WPI; 2004-347921/32.	
XX	N-PSDB; ACN40141.	
PT	New tumor-associated antigenic target polypeptides and nucleic acids,	
PT	useful in preparing a medicament for treating or detecting a	
PT	proliferative disorder, e.g. breast, lung, colorectal, ovarian or	
PT	prostate cancer or tumor.	
PS	Claim 12; SEQ ID NO 4738; 7273BP; English.	
CC	The invention relates to human tumour-associated antigenic target (TAT)	
CC	polypeptides, and their related nucleic acids. The TAT polypeptides are	
CC	overexpressed in cancer tissues compared to normal tissues, and may thus	
CC	serve as effective targets for the diagnosis and treatment of cancer in	
CC	mammals. The invention also relates to nucleic acid and polypeptide	
CC	sequences at least 80% identical to the TAT nucleic acids and	
CC	polypeptides; expression vectors and host cells comprising a TAT nucleic	
CC	acid; an antibody specific for a TAT polypeptide; a peptide or organic	
CC	molecule which binds to a TAT polypeptide; fusion proteins comprising a	
CC	TAT polypeptide; and methods and compositions for the treatment or	
CC	diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,	
CC	antibodies, antagonists, binding molecules and compositions are useful	
CC	for diagnosing or treating a cell proliferative disorder associated with	
CC	increased TAT expression, particularly cancers such as breast cancer,	
CC	colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder	
CC	cancer, pancreatic cancer, cervical cancer, cancers of the central	
CC	nervous system, melanoma and leukaemia. TAT nucleic acids may further be	
CC	used as hybridisation probes, in chromosome and gene mapping, in	
CC	chromosome identification and in gene therapy. The present sequence	

CC	represents a TAT polypeptide of the invention					
XX						
SQ	Sequence	146 AA;				
	Query Match	1.3%;	Score 8;	DB 8;	Length 146;	
	Best Local Similarity	100.0%;	Pred.	No. 1.2e+02;		
	Matches	8;	Conservative	0;	Mismatches	0; Indels
Oy	234 EQQQQQQLQ	241				
Db	112 EQQQQQQLQ	119				
<hr/>						
RESULT 59						
AAG32017						
ID	AAG32017 standard; protein; 167 AA.					
XX						
AC	AAG32017;					
DT	17-OCT-2000 (first entry)					
XX						
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 38547.					
XX						
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.					
XX						
OS	Arabidopsis thaliana.					
PN	EP1033405-A2.					
XX						
PD	06-SEP-2000.					
XX						
PF	25-FEB-2000; 2000EP-00301439.					
XX						
PR	25-FEB-1999;	99US-0121825P.				
PR	05-MAR-1999;	99US-0123160P.				
PR	09-MAR-1999;	99US-0123548P.				
PR	23-MAR-1999;	99US-0125788P.				
PR	25-MAR-1999;	99US-0126264P.				
PR	29-MAR-1999;	99US-0126785P.				
PR	01-APR-1999;	99US-0127462P.				
PR	06-APR-1999;	99US-0128234P.				
PR	08-APR-1999;	99US-0128714P.				
PR	16-APR-1999;	99US-0129845P.				
PR	19-APR-1999;	99US-0130077P.				
PR	21-APR-1999;	99US-0130449P.				
PR	23-APR-1999;	99US-0130510P.				
PR	23-APR-1999;	99US-0130891P.				
PR	28-APR-1999;	99US-0131449P.				
PR	30-APR-1999;	99US-0132048P.				
PR	30-APR-1999;	99US-0132407P.				
PR	04-MAY-1999;	99US-0132484P.				
PR	05-MAY-1999;	99US-0132485P.				
PR	06-MAY-1999;	99US-0132486P.				
PR	06-MAY-1999;	99US-0132487P.				
PR	07-MAY-1999;	99US-0132863P.				
PR	11-MAY-1999;	99US-0134256P.				
PR	14-MAY-1999;	99US-0134218P.				
PR	14-MAY-1999;	99US-0134219P.				
PR	14-MAY-1999;	99US-0134221P.				
PR	14-MAY-1999;	99US-0134370P.				
PR	18-MAY-1999;	99US-0134768P.				
PR	19-MAY-1999;	99US-0134941P.				
PR	20-MAY-1999;	99US-0135124P.				
PR	21-MAY-1999;	99US-0135353P.				
PR	25-MAY-1999;	99US-0135629P.				
PR	25-MAY-1999;	99US-0136021P.				
PR	27-MAY-1999;	99US-0136392P.				
PR	28-MAY-1999;	99US-0136782P.				
PR	01-JUN-1999;	99US-0137222P.				
PR	03-JUN-1999;	99US-0137528P.				
PR	04-JUN-1999;	99US-0137502P.				

PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.

PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148311P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158222P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160815P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.3%; Score 8; DB 3; Length 167;
Best local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGSGA 215
 Db 55 GGGGSGA 62

RESULT 60
 ID ABO61854 standard; protein; 169 AA.
 AC ABO61854;
 DT 29-JUL-2004 (first entry)
 XX Klebsiella pneumoniae polypeptide seqid 8371.
 XX
 KM Recombinant expression vector; transcription regulatory element;
 KM Klebsiella pneumoniae protein; antibacterial; Vaccine.
 XX Klebsiella pneumoniae.
 OS
 PN US6610836-B1.
 PD 26-AUG-2003.
 PF 27-JAN-2000; 2000US-00489039.
 XX
 PR 29-JAN-1999; 99US-0117747P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Breton GU, Osborne M;
 XX WPI; 2003-895346/82.
 DR N-PSDB; ACH95405.
 XX
 PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
 preparing a vaccine composition against Klebsiella pneumoniae.
 XX
 PS Disclosure; SEQ ID NO 8371; 932P; English.
 XX
 CC The invention describes a new isolated nucleic acid encoding a Klebsiella
 CC pneumoniae polypeptide. Also described are: a recombinant expression
 CC vector comprising the nucleic acid, operably linked to a transcription
 CC regulatory element; and a cell comprising the recombinant expression
 CC vector. The nucleic acid is useful for preparing a vaccine composition
 CC against Klebsiella pneumoniae. This is the amino acid sequence of a
 CC Klebsiella pneumoniae polypeptide of the invention
 XX
 SQ Sequence 169 AA;

Query Match 1.3%; Score 8; DB 7; Length 169;
 Best Local Similarity 100.0%; Pred.No.1.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 PPSPLPL 446
 Db 39 PPSPLPL 46

RESULT 61
 ID AAG54374 standard; protein; 171 AA.
 AC AAG54374;
 DT 18-OCT-2000 (first entry)
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 69323.
 DE
 XX Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 KM termination sequence.

XX Arabidopsis thaliana.
 OS
 XX EPI033405-A2.
 PN
 XX 06-SEP-2000.
 PD
 XX
 PF 25-FEB-2000; 2000EP-00301439.
 XX
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128334P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.
 PR 28-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 30-APR-1999; 99US-0132407P.
 PR 04-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.
 PR 10-JUN-1999; 99US-0138540P.
 PR 10-JUN-1999; 99US-0138847P.
 PR 14-JUN-1999; 99US-0139119P.
 PR 16-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.
 PR 28-JUN-1999; 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142803P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150864P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.

PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156569P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 13-OCT-1999; 99US-0159296P.
PR 14-OCT-1999; 99US-0159339P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.3%; Score 8; DB 3; Length 171;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GCGGSGA 215
Db 75 GCGGSGA 82

RESULT 62
AAG42727
ID AAG42727 standard; protein; 171 AA.

AC AAG42727;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 53318.
XX DE
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX

PF 25-FEB-2000; 2000BP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131499P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135333P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.

PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147433P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 22-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-015138P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.

PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158032P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161982P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.3%; Score 8; DB 3; Length 171;
Best Local Similarity 100.0%; Pred.No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GGGGGSGA 215
|||||
Db 75 GGGGGSGA 82

RESULT 63

AGG23183
ID AAG23183 standard; protein; 171 AA.

AC AAG23183;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 26394.

XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

XX Arabidopsis thaliana.

OS EP1033405-A2.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129645P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136382P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137528P.
PR 03-JUN-1999; 99US-0137529P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137742P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138547P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139482P.
PR 18-JUN-1999; 99US-0139484P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139464P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 24-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.

PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145267P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147303P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148317P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149466P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151458P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.

PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158929P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161049P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161820P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.3%; Score 8; DB 3; Length 171;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

Oy 208 GGGGSSGA 215
Db 75 GGGGSSGA 82

RESULT 64
ABU06773
ID ABU06773 standard; protein; 176 AA.
XX
AC ABU06773;
XX
DT 17-FEB-2003 (first entry)
XX
DE Maize starch synthase IIB (SSIIB) associated protein #1.
XX
KW Starch; starch synthase; glucan association domain; GLASS; linker domain;
KW LINKR; glucosyl transferase domain; GLYTR; C-terminal end; CTEND;
KW granule bound starch synthase; GBSS; morphology; retrogradation;
KW waterbinding; swelling potential.
XX
OS Unidentified.
XX
PN WO200279410-A2.
XX
PD 10-OCT-2002.
XX
PF 29-MAR-2002; 2002WO-US009574.
XX
PR 30-MAR-2001; 2001US-0279720P.
XX
PA (BADI) BASF PLANT SCI GMBH.
XX
PI Commuri P, Keeling PJ, Ramirez N, McKean A, Gao Z, Guan H;
XX WPI; 2003-040678/03.
DR
XX
PT New DNA encoding fusion protein consisting of 4 different functional
PT domains selected from glucan association domain, linker domain, glucosyl
PT transferase domain, and C-terminal end, useful for producing modified

CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences AB067826-
CC AB084396 represent P. aeruginosa polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 184 AA;

Query Match 1.3%; Score 8; DB 7; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GSEPAAG 122
Db 36 GSEPAAG 43

RESULT 67
AAG32016
ID AAG32016 standard; protein; 187 AA.
XX
AC AAG32016;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 38546.
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134421P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140035P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141087P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0143452P.
PR 13-JUL-1999; 99US-0143624P.
PR 14-JUL-1999; 99US-0144005P.
PR 15-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.

PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 06-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 09-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 16-AUG-1999; 99US-0148684P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153756P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 26-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.

PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161982P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query March 1.3%; Score 8; DB 3; Length 187;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Gaps 0;
 Indels 0;

Dy 208 GCGGSGGA 215
 Db 75 GCGGSGGA 82

RESULT 68
 ABB21687
 ID ABB21687 standard; protein; 191 AA.
 AC ABB21687;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Protein #3686 encoded by probe for measuring heart cell gene expression.
 XX
 KW Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 XX
 OS Homo sapiens.
 XX
 FN W0200157274-A2.
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001MO-US000666.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLR-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488899/53.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts.
 XX
 PS Claim 15; SEQ ID NO 23457; 530bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC AB21335-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 191 AA;

Query Match 1.3%; Score 8; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LGGGGSG 214
DB 82 LGGGGSG 89

RESULT 69
ID AAM69485 standard; protein; 191 AA.
AC AAM69485;
XX
XX
XX AAM69485;
XX
XX
XX 06-NOV-2001 (first entry)
XX
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 29791.
XX
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000668.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX
XX
XX Example 4; SEQ ID NO 29791; 658bp + Sequence Listing; English.
XX
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX
XX
XX Sequence 191 AA;

Query Match 1.3%; Score 8; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LGGGGSG 214
DB 82 LGGGGSG 89

RESULT 70
ID AAM57094 standard; protein; 191 AA.
XX

AC AAM57094;
XX
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 29199.
XX
XX
XX
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000667.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brain.
XX
XX Example 4; SEQ ID NO 29199; 650bp + Sequence Listing; English.
XX
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX
XX
XX Sequence 191 AA;

Query Match 1.3%; Score 8; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LGGGGSG 214
DB 82 LGGGGSG 89

RESULT 71
ID AAM05009 standard; protein; 191 AA.
XX
XX
XX AAM05009;
XX
XX
XX 09-OCT-2001 (first entry)
XX
XX
XX Peptide #3691 encoded by probe for measuring breast gene expression.
XX
XX Probe; human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX

XX PF 29-JAN-2001; 2001WO-US000661.
 XX PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-0063236P.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PT a human breast.
 XX DR WPI; 2001-476286/51.
 XX PT Novel single exon nucleic acid probe used to measuring gene expression in
 PT a human breast.
 XX PS Claim 27; SEQ ID NO 13749; 322pp; English.
 XX CC The present invention relates to novel single exon nucleic acid probes
 CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridizes at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of development,
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC XX
 SQ Sequence 191 AA;
 Query Match 1.3%; Score 8; DB 4; Length 191;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 207 LGGGGGSG 214
 Db 82 LGGGGGSG 89
 RESULT 72
 ADG27706
 ID ADG27706 standard; protein; 191 AA.
 XX AC ADG27706;
 XX DT 26-FEB-2004 (first entry)
 XX DE Human novel protein amino acid sequence SeqID473.
 XX KW antiparkinsonian; haemostatic; nootropic; neuroprotective; osteopathic;
 KW anti-HIV; protozoacide; antifungal; immunosuppressive; antirheumatic;
 KW antiarthritic; antidiabetic; antiallergic; antiinflammatory;
 KW anticoagulant; cytostatic; gene therapy; Parkinson's disease;
 KW Alzheimer's disease; thrombocytopaenia; osteoporosis; osteoarthritis;
 KW infection; HIV; leishmania; malaria; fungal infection;
 KW multiple sclerosis; rheumatoid arthritis;
 KW insulin dependent diabetes mellitus; allergic reaction; food allergy;
 KW insect allergy; allergic rhinitis; haemophilia; cancer; human.
 XX OS Homo sapiens.
 XX PN WO200179254-A1.
 XX PD 25-OCT-2001.

PF 16-APR-2001; 2001WO-US008655.
 XX PR 18-APR-2000; 2000US-00552929.
 PR 22-SEP-2000; 2000US-00668317.
 PR 24-OCT-2000; 2000US-00695783.
 PR 01-DEC-2000; 2000US-00728628.
 PR 26-JAN-2001; 2001US-00770160.
 PR 13-FEB-2001; 2001US-00783066.
 PR 22-MAR-2001; 2001US-00816828.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J;
 PI Yang Y, Zhao QA, Goodrich RW, Liu C, Drmanac RT, Ma Y, Wang Z;
 PI Wehrman T;
 XX DR WPI; 2001-607699/69.
 XX PT Novel polynucleotides and encoded polypeptides (protein factors,
 PT including, e.g., cytokines, such as lymphokines, interferons, and
 PT circulating soluble factors) useful for treating, e.g., Parkinson's,
 PT Alzheimer's, HIV and cancer.
 XX PS Claim 20; SEQ ID NO 473; 153pp; English.
 XX CC This invention relates to a novel isolated DNA sequence and the mature
 CC proteins encoded by them. The invention may be useful in the development
 CC of compositions with antiparkinsonian, haemostatic, nootropic,
 CC neuroprotective, osteopathic, anti-HIV, protozoacide, antifungal,
 CC immunosuppressive, antirheumatic, antiarthritic, antidiabetic,
 CC anti-allergic, antiinflammatory, anticoagulant or cytostatic activities.
 CC In addition, the sequences of the invention may be useful for gene
 CC therapy. The invention may be useful for the development of treatments
 CC for Parkinson's, Alzheimer's, thrombocytopaenia, osteoporosis,
 CC osteoarthritis, infections (including HIV, leishmania, malaria, and
 CC various fungal infection), autoimmune disorders such as multiple
 CC sclerosis, rheumatoid arthritis, and insulin dependent diabetes mellitus,
 CC allergic reactions and conditions (for example food allergies, insect
 CC allergies and allergic rhinitis), coagulation disorders including
 CC haemophilia, and cancer. Note: The amino acid sequences given in table 6
 CC (SeqID 439-584) may have in frame stop codons or possible
 CC insertions/deletions as shown in the table. The sequences allocated Seq
 CC IDs 1-438 are not provided (even by reference) in the specification.
 CC XX
 SQ Sequence 191 AA;
 Query Match 1.3%; Score 8; DB 4; Length 191;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 66 SPAPSPAG 73
 Db 141 SPAPSPAG 148
 RESULT 73
 AD065667
 ID AD065667 standard; protein; 202 AA.
 XX AC AD065667;
 XX DT 07-OCT-2004 (first entry)
 XX DE Novel human protein sequence #640.
 XX KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
 KW gene therapy; diagnostic marker; morbid state; osteoporosis;
 KW neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
 KW cancer.
 XX OS Homo sapiens.
 XX PN EP1440981-A2.

```

XX 28-JUL-2004.
PD
XX
XX 21-JAN-2004; 2004BP-00001196.
PF
XX
XX 21-JAN-2003; 2003JP-00102206.
PR
XX 09-MAY-2003; 2003JP-00131392.
PA
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Nagai K, Irie R;
XX MPI; 2004-535376/52.
DR N-PSDB; AD063479.
XX
XX Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
XX Claim 1; SEQ ID NO 2828; 2449pp; English.
PS
XX The invention relates to 2495 novel polynucleotides (I) and their encoded
CC polypeptides, sequences hybridizing to these nucleotides, sequences
CC encoding partial polypeptides and sequences having 70% or 90% identity to
CC the nucleotide and protein sequences. The nucleotides and polypeptides
CC are useful as diagnostic markers or therapeutic target for the diseases
CC or morbid states. They are also useful for treating osteoporosis,
CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC dementia and various cancers. This sequence corresponds to a protein
CC sequence of the invention.
XX
XX Sequence 202 AA;
SQ
XX
XX Query Match 1.3%; Score 8; DB 8; Length 202;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 439 PPPSPPL 446
DB 78 PPPSPPL 85
XX
XX RESULT 74
XX ABO60152
ID ABO60152 standard; protein; 208 AA.
XX
XX ABO60152;
AC
XX 29-JUL-2004 (first entry)
DT
XX Human genome derived single exon protein #6386.
XX
XX Human; gene expression; single exon probe; microarray;
KM alternative splicing event; genomic alteration.
XX
XX Homo sapiens.
OS
XX US2003194704-A1.
PN
XX 16-OCT-2003.
PD
XX 03-APR-2002; 2002US-00029386.
XX
XX 03-APR-2002; 2002US-00029386.
PR
XX (PENN/) PENN S G.
PA (RANK/) RANK D R.
XX (HANZ/) HANZEL D K.
XX
XX Penn SG, Rank DR, Hanzel DK,
PI
XX MPI; 2004-119264/12.
XX

```

```

PT New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for
XX surveying tissues.
XX
XX Claim 45; SEQ ID NO 33786; 80pp; English.
PS
XX The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule
CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above-mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC methods of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC human gene expression data by subcription, and a computer-readable
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above. The probe, methods and apparatus are useful in gene
CC expression analysis. The probes may be used as tools for surveying
CC tissues to detect the presence of expressed messages that contain their
CC specific exon, or in constructing genome-derived single exon microarrays.
CC In addition, the probes are used in identifying and characterising
CC alternative splicing events, in detecting and characterising gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe protein of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docID=20030194704
XX
XX Sequence 208 AA;
SQ
XX
XX Query Match 1.3%; Score 8; DB 8; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 207 LGGGGSSG 214
DB 82 LGGGGSSG 89
XX
XX RESULT 75
XX ABB67341
ID ABB67341 standard; protein; 216 AA.
XX
XX ABB67341;
AC
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 28815.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX W0200171042-A2.
PN
XX 27-SEP-2001.
XX

```

PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL11444.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 28815; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 216 AA;
 Query Match 1.3%; Score 8; DB 4; Length 216;
 Best local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 32 GGPEDEPG 39
 |||||
 Db 74 GGPEDEPG 81

Search completed: March 23, 2005, 15:48:53
 Job time : 177 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2005, 11:49:41 ; Search time 2396 Seconds
(without alignments)
10874.040 Million cell updates/sec

Title: US-10-054-935-1
Sequence: 1 cagctcgcaccccccgcac.....aaaaaaaaaaaaaaaa 4372

Scoring table: IDENTITY_NIC
Gapop 10.0, Gapext 1.0

Searched: 5552208 seqs, 297965951 residues
Total number of hits satisfying chosen parameters: 11104416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Published Applications NA:
2: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:
3: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:
4: /cgn2_6/prodata/1/pubpna/US05_PUBCOMB.seq:
5: /cgn2_6/prodata/1/pubpna/US04_PUBCOMB.seq:
6: /cgn2_6/prodata/1/pubpna/US03_PUBCOMB.seq:
7: /cgn2_6/prodata/1/pubpna/US02_PUBCOMB.seq:
8: /cgn2_6/prodata/1/pubpna/US01_PUBCOMB.seq:
9: /cgn2_6/prodata/1/pubpna/US00_PUBCOMB.seq:
10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:
12: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:
13: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:
14: /cgn2_6/prodata/1/pubpna/US05_PUBCOMB.seq:
15: /cgn2_6/prodata/1/pubpna/US04_PUBCOMB.seq:
16: /cgn2_6/prodata/1/pubpna/US03_PUBCOMB.seq:
17: /cgn2_6/prodata/1/pubpna/US02_PUBCOMB.seq:
18: /cgn2_6/prodata/1/pubpna/US01_PUBCOMB.seq:
19: /cgn2_6/prodata/1/pubpna/US00_PUBCOMB.seq:
20: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:
21: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:
22: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	4372	100.0	4372	US-10-054-935-1	Sequence 1, Appl
2	4372	100.0	4372	US-10-144-194A-65	Sequence 65, Appl
3	4372	100.0	4372	US-10-491-566-65	Sequence 65, Appl
4	1508.4	34.5	1545	US-10-641-643-167	Sequence 167, Appl
5	1478	33.8	1509	US-09-397-945-79	Sequence 79, Appl
6	1478	33.8	1509	US-10-653-595-79	Sequence 79, Appl
7	837	19.1	1529	US-09-925-102-212	Sequence 212, Appl
8	837	19.1	1529	US-09-925-102-212	Sequence 212, Appl
9	596	13.6	1290	US-10-641-643-322	Sequence 322, Appl
10	565.8	12.9	600	US-10-242-535A-51339	Sequence 51339, A
11	565.8	12.9	600	US-10-085-783A-51339	Sequence 51339, A

12	551.6	12.6	615	9	US-09-917-800A-725	Sequence 725, App
13	510.2	11.7	542	9	US-09-736-457-1554	Sequence 1554, App
14	510.2	11.7	542	9	US-09-902-941-1554	Sequence 1554, App
15	510.2	11.7	542	9	US-09-849-626-1554	Sequence 1554, App
16	510.2	11.7	542	14	US-10-017-754-1554	Sequence 1554, App
17	510.2	11.7	542	16	US-10-113-872-1554	Sequence 1554, App
18	510.2	11.7	542	17	US-10-283-017-1554	Sequence 1554, App
19	448	10.2	451	17	US-10-242-535A-58053	Sequence 58053, A
20	448	10.2	451	17	US-10-085-783A-58053	Sequence 58053, A
21	389.6	8.9	365	10	US-09-918-995-30064	Sequence 30064, A
22	357	8.2	365	10	US-09-918-995-30064	Sequence 8833, App
23	320.2	7.3	400	17	US-10-085-783A-8833	Sequence 8833, App
24	320.2	7.3	400	17	US-10-085-783A-8833	Sequence 8833, App
25	274.8	6.3	590	9	US-09-917-800A-16	Sequence 16, Appl
26	274.8	6.3	590	17	US-10-191-800A-451	Sequence 451, Appl
27	263	6.0	273	18	US-10-425-115-150265	Sequence 150265, A
28	243.4	5.6	245	17	US-10-242-535A-43646	Sequence 43646, A
29	243.4	5.6	245	17	US-10-085-783A-43646	Sequence 43646, A
30	230	5.3	230	9	US-09-983-965-2328	Sequence 2127, App
31	228	5.2	346	9	US-09-983-965-2328	Sequence 2127, App
32	214	4.9	440	17	US-10-242-535A-31928	Sequence 31928, A
33	214	4.9	440	17	US-10-085-783A-31928	Sequence 31928, A
34	176.2	4.0	196	15	US-10-106-698-3799	Sequence 3799, App
35	134.8	3.1	211	18	US-10-357-930-56465	Sequence 56465, A
36	107.2	2.5	565	9	US-09-917-800A-737	Sequence 737, App
37	74.4	1.7	763	18	US-10-437-963-49370	Sequence 49370, A
38	73.4	1.7	972	18	US-10-425-115-17653	Sequence 17653, A
39	71.4	1.6	4667	18	US-10-723-860-5759	Sequence 5759, App
40	70.4	1.6	723	10	US-09-814-353-4726	Sequence 4726, App
41	70.4	1.6	723	10	US-09-814-353-4726	Sequence 4726, App
42	67.6	1.5	433	18	US-10-814-323-6681	Sequence 11024, A
43	66.8	1.5	488	10	US-09-918-995-15186	Sequence 6681, App
44	66.8	1.5	488	10	US-09-918-995-15186	Sequence 15186, A
45	66.8	1.5	719	18	US-10-425-115-74544	Sequence 74544, A

ALIGNMENTS

RESULT 1
US-10-054-935-1
Sequence 1, Application US/10054935
Publication No. US2003013546A1
GENERAL INFORMATION:
APPLICANT: Origene Technologies, Inc
TITLE OF INVENTION: BREAST CANCER TRANSCRIPTION-FACTOR GENE AND USES
FILE REFERENCE: 16U 107 R1
CURRENT APPLICATION NUMBER: US/10/054, 935
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 4372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (78)..(1922)
OTHER INFORMATION:
US-10-054-935-1

Query Match	100.0%	Score 4372;	DB 15;	Length 4372;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 4372;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CAGTCTCGACCCCGACCTCGCCCTTCCACCCCTCTCTCGCTGCGCGG	60	
Db	1	CAGTCTCGACCCCGACCTCGCCCTTCCACCCCTCTCTCGCTGCGCGG	60	
Qy	61	CGCTCTCGGACACATATGATGATCGCGGTTCAGCGCGCGCGCCCTG	120	
Db	61	CGCTCTCGGACACATATGATGATCGCGGTTCAGCGCGCGCGCCCTG	120	

Sum
Li
Davi
Korec

QY 122 CGGCGGCAATCCTGAGCAGCGACTGACAGAGCGGGCTGCGCGCTGGGCGGCGCG 180
Db 121 CGGCGGCAATCCTGAGCAGCGACTGACAGAGCGGGCTGCGCGCTGGGCGGCGCG 180
QY 181 AGGACGAGCTGGGGCGGCGGAGGCCACTTCTCCCGGCAACCGTAAGCTCAAGAGC 240
Db 181 AGGACGAGCTGGGGCGGCGGAGGCCACTTCTCCCGGCAACCGTAAGCTCAAGAGC 240
QY 241 CGGGGCGCGGCTGGCGCTCTCCAGAGGCGGAGCGCGGCTTCCCGGCGGCTGCG 300
Db 241 CGGGGCGCGGCTGGCGCTCTCCAGAGGCGGAGCGCGGCTTCCCGGCGGCTGCG 300
QY 301 GCGGCAAGGCGCGGCTTGTATCTCCGCGCGGGGCGCGCGCGGCAACAGAGAGAG 360
Db 301 GCGGCAAGGCGCGGCTTGTATCTCCGCGCGGGGCGCGCGCGGCAACAGAGAGAG 360
QY 361 GCTGGGCGGCTTGGTGCCTTTCCTGTCGCGCGCGGCAACAGAGAGAGAGAG 420
Db 361 GCTGGGCGGCTTGGTGCCTTTCCTGTCGCGCGCGGCAACAGAGAGAGAGAG 420
QY 421 GGGGAGAGCTGCGCAGCGGAGCGGCTGAGCGCGCGCGCGCAAGTATCAGGCGTGC 480
Db 421 GGGGAGAGCTGCGCAGCGGAGCGGCTGAGCGCGCGCGCGCAAGTATCAGGCGTGC 480
QY 481 TGGCCATTGAGAGGGGCTCTGTGGCGGCGCGCAAGAGCCTTACGCGCTGGGCTGGG 540
Db 481 TGGCCATTGAGAGGGGCTCTGTGGCGGCGCGCAAGAGCCTTACGCGCTGGGCTGGG 540
QY 541 ACAAGGATGGGGCGGCTCCCGCGCTGCGCAGCGGCTGAGCGCGGAGCGCGCAC 600
Db 541 ACAAGGATGGGGCGGCTCCCGCGCTGCGCAGCGGCTGAGCGCGGAGCGCGCAC 600
QY 601 TACCTTGCCCGGCGCGCACCCCTGCGCGCACCGCGCACCGCGGAGCCTTGCGGCGCA 660
Db 601 TACCTTGCCCGGCGCGCACCCCTGCGCGCACCGCGCACCGCGGAGCCTTGCGGCGCA 660
QY 661 GCGAGGCGAGTGGAGAGATGAGAGAGAGCGCTTCGCGGGGATGGGCGGCTCGGAG 720
Db 661 GCGAGGCGAGTGGAGAGATGAGAGAGAGCGCTTCGCGGGGATGGGCGGCTCGGAG 720
QY 721 CCTCCAGTCAGGCGCGCTGCTCAACAGATCTTCTGCTGCAATTGAGCCTCATCGAAC 780
Db 721 CCTCCAGTCAGGCGCGCTGCTCAACAGATCTTCTGCTGCAATTGAGCCTCATCGAAC 780
QY 781 AGCAGCAGCAGCTGCGAGCGCAAGAGAGAGATGAGAGCTGAGACTCAGAGAG 840
Db 781 AGCAGCAGCAGCTGCGAGCGCAAGAGAGAGATGAGAGCTGAGACTCAGAGAG 840
QY 841 ACACGCTCTGCTCGGATTTGAACTGTAAGAGGCGGATGCGCTGTTAAAGAGATA 900
Db 841 ACACGCTCTGCTCGGATTTGAACTGTAAGAGGCGGATGCGCTGTTAAAGAGATA 900
QY 901 ACAGAGAGAGAGCAGAGCTTTTCAAGGGCTATGAACTGAGAGAGAGAGAGAG 960
Db 901 ACAGAGAGAGAGCAGAGCTTTTCAAGGGCTATGAACTGAGAGAGAGAGAGAG 960
QY 961 AGCTATCTGAGAGAAATTTAACTGAGTGCAGCGGAGCTTTTCGAGACATCCCAAGCTC 1020
Db 961 AGCTATCTGAGAGAAATTTAACTGAGTGCAGCGGAGCTTTTCGAGACATCCCAAGCTC 1020
QY 1021 TGGCTCCCAAGCGCTTCTCATGTGGGCGGAGTGAAAGGAGCATTAAGAGAAATCCCAT 1080
Db 1021 TGGCTCCCAAGCGCTTCTCATGTGGGCGGAGTGAAAGGAGCATTAAGAGAAATCCCAT 1080
QY 1081 TTGGAAGTACAGAAAGAGACTCTGTATAAAGCTGCTCTGTAATTTTCAAAAGTCA 1140
Db 1081 TTGGAAGTACAGAAAGAGACTCTGTATAAAGCTGCTCTGTAATTTTCAAAAGTCA 1140
QY 1141 AAAAGAGAGCTCTTAAGCACTCTCTTAATAAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db 1141 AAAAGAGAGCTCTTAAGCACTCTCTTAATAAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 CTGTATTGAAGCTGAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

Db 1201 CTGTATTGAAGCTGAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 ACAAGGAGAGAGCTCTCACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1261 ACAAGGAGAGAGCTCTCACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 CTTTCTCAAGTGAATGAAGATTTGCGTACTTTTCAACACAGAGAGAGAGAGAGAG 1380
Db 1321 CTTTCTCAAGTGAATGAAGATTTGCGTACTTTTCAACACAGAGAGAGAGAGAGAG 1380
QY 1381 GTTGGAGAGAGAGCTCCCATCAAGCTTACATTAAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 GTTGGAGAGAGAGCTCCCATCAAGCTTACATTAAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 AGAGTGAAGAGAGTCTGATGAGCATCAAGTGTTCAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1441 AGAGTGAAGAGAGTCTGATGAGCATCAAGTGTTCAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 TTTCTTTCTGAG 1560
Db 1501 TTTCTTTCTGAG 1560
QY 1561 TGGAG 1620
Db 1561 TGGAG 1620
QY 1621 AG 1680
Db 1621 AG 1680
QY 1681 AGCTCAGATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Db 1681 AGCTCAGATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1741 TCCCTGAG 1800
Db 1741 TCCCTGAG 1800
QY 1801 TTGAG 1860
Db 1801 TTGAG 1860
QY 1861 GTAGCGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 1861 GTAGCGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1921 AGCTGAG 1980
Db 1921 AGCTGAG 1980
QY 1981 AG 2040
Db 1981 AG 2040
QY 2041 TCACTACTCTGAG 2100
Db 2041 TCACTACTCTGAG 2100
QY 2101 AATGTGAG 2160
Db 2101 AATGTGAG 2160
QY 2161 AAGCTAAG 2220
Db 2161 AAGCTAAG 2220
QY 2221 CTAATGTAAG 2280
Db 2221 CTAATGTAAG 2280
QY 2281 TCCCTTGAAG 2340

D 2281 TCCCTTAAGACCTGCTGCATGCGACTCTTCTCACCCCTGGTACCCCTCTTAATAGTGG 2340
Q 2341 GATATAGATATTTTAAACCTTAATAATAAACAACCTCAGCATAGACTTTAGAACAG 2400
D 2341 GATATAGATATTTTAAACCTTAATAATAAACAACCTCAGCATAGACTTTAGAACAG 2400
Q 2401 AAGAAGATGACAAAGTGAAGCGATGAAGCAAGCCATCTTCAAGAGTGAAGAAAGACATCG 2460
D 2401 AAGAAGATGACAAAGTGAAGCGATGAAGCAAGCCATCTTCAAGAGTGAAGAAAGACATCG 2460
Q 2461 GAGAGTGGATGATATGCTGTAAGAAAGATAGTTGTTCACTTGAATAATATCTGTGTAC 2520
D 2461 GAGAGTGGATGATATGCTGTAAGAAAGATAGTTGTTCACTTGAATAATATCTGTGTAC 2520
Q 2521 AGTCATGAGGAAGGAGATGTTGGCTGTGATATTTTTCAGTTAATGATTAACAATTTTC 2580
D 2521 AGTCATGAGGAAGGAGATGTTGGCTGTGATATTTTTCAGTTAATGATTAACAATTTTC 2580
Q 2581 TTTTACTGCTCAAAAACCAAAATCTTTGGAAAAGAAAGTGGGGATGTTAGTTTCAGACA 2640
D 2581 TTTTACTGCTCAAAAACCAAAATCTTTGGAAAAGAAAGTGGGGATGTTAGTTTCAGACA 2640
Q 2641 AGTTACAGCTGTAAACAAAGCACTTATGTTGGATGGCATGGCCAAAACCTGTATAAA 2700
D 2641 AGTTACAGCTGTAAACAAAGCACTTATGTTGGATGGCATGGCCAAAACCTGTATAAA 2700
Q 2701 TGTCTGTATGATCACTCTTCAAGTATCTTCAATGGGCTTCACTCTTTTTCAGAGA 2760
D 2701 TGTCTGTATGATCACTCTTCAAGTATCTTCAATGGGCTTCACTCTTTTTCAGAGA 2760
Q 2761 ACTCTTGGTGGATGATGAGACTTGAAGAGGATGGGGAGAGTGTGAATAATAGTGTCT 2820
D 2761 ACTCTTGGTGGATGATGAGACTTGAAGAGGATGGGGAGAGTGTGTGAATAATAGTGTCT 2820
Q 2821 TCCCTTGGCTGGCAAAATGTCTAATCTTGAACAAAGATGTAATGAGCTTTCTCC 2880
D 2821 TCCCTTGGCTGGCAAAATGTCTAATCTTGAACAAAGATGTAATGAGCTTTCTCC 2880
Q 2881 ATTCACTTGTAAATAATTTTGTATGTATGATCACTTGTGGCTCTCCCTCCCGCTTTTG 2940
D 2881 ATTCACTTGTAAATAATTTTGTATGTATGATCACTTGTGGCTCTCTCCCTCCCGCTTTTG 2940
Q 2941 TTAATAATATCAGATATGACCTCCAGGCACTTGTGTCTCAGTGTAAAGTCCCTATTAC 3000
D 2941 TTAATAATATCAGATATGACCTCCAGGCACTTGTGTCTCAGTGTAAAGTCCCTATTAC 3000
Q 3001 TATCTGAAGGAATAATGAGCCAAAGACCTGTGTCTCAAAATATATGAAATTGCTTTCT 3060
D 3001 TATCTGAAGGAATAATGAGCCAAAGACCTGTGTCTCAAAATATATGAAATTGCTTTCT 3060
Q 3061 TTAGCTTTCAGACTATTTGTGTGAACAAGTAGGGGCTTAATCTCTAGAAAGTAGGGG 3120
D 3061 TTAGCTTTCAGACTATTTGTGTGAACAAGTAGGGGCTTAATCTCTAGAAAGTAGGGG 3120
Q 3121 CTTTATATCTTAAAGAGATATGCTCCAGATTTATAGCACTTTTAAAGAGAACCAAG 3180
D 3121 CTTTATATCTTAAAGAGATATGCTCCAGATTTATAGCACTTTTAAAGAGAACCAAG 3180
Q 3181 GATATGAGGTGTGTGCTGAGCCATCAGTGAAGCAGAAAGAGAAATGAGATCAATTC 3240
D 3181 GATATGAGGTGTGTGCTGAGCCATCAGTGAAGCAGAAAGAGAAATGAGATCAATTC 3240
Q 3241 TGGGAAGAAGAAAGTTCCTCAGGGGCTCCCACTGCTAAATTTTGTGTGAGTGT 3300
D 3241 TGGGAAGAAGAAAGTTCCTCAGGGGCTCCCACTGCTAAATTTTGTGTGAGTGT 3300
Q 3301 GATCTGTGCTTCTGATTTGATCTTTTAAAGAAATTTTTCGACAGACATGATATTC 3360
D 3301 GATCTGTGCTTCTGATTTGATCTTTTAAAGAAATTTTTCGACAGACATGATATTC 3360
Q 3361 TTGGATATCTTGTGCTCTTATTTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGT 3420
D 3361 TTGGATATCTTGTGCTCTTATTTTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGT 3420

Q 3421 TGGGTTTCAATTTGTAATCTCATCTGCTTGAAGAGTGGGCTCTTAATAAGGAACCTGC 3480
D 3421 TGGGTTTCAATTTGTAATCTCATCTGCTTGAAGAGTGGGCTCTTAATAAGGAACCTGC 3480
Q 3481 TGTAACTTCATTTGACGCAAGGATGTGAAGAAATGAGACTTAATTTCCACTGAGGGCTC 3540
D 3481 TGTAACTTCATTTGACGCAAGGATGTGAAGAAATGAGACTTAATTTCCACTGAGGGCTC 3540
Q 3541 TCACTCACACTTAAGAGAGATTTCTGAAGAACTGGGCCAGATTTTCTTGTGCTC 3600
D 3541 TCACTCACACTTAAGAGAGATTTCTGAAGAACTGGGCCAGATTTTCTTGTGCTC 3600
Q 3601 CATCATTTAATGTGACAGGCTGTTCAATTTCTTACTCTTAATGTATATTTCTTC 3660
D 3601 CATCATTTAATGTGACAGGCTGTTCAATTTCTTACTCTTAATGTATATTTCTTC 3660
Q 3661 GATACGTGTCCAAAAGAAAAAGAACCCAAATCAGTGTCTTGAATTTTGTGTTGATTC 3720
D 3661 GATACGTGTCCAAAAGAAAAAGAACCCAAATCAGTGTCTTGAATTTTGTGTTGATTC 3720
Q 3721 CTCAGTTTCTCTGATTTGAGATGTGTGGGCTCTAATTTTGGGTATGATGACAA 3780
D 3721 CTCAGTTTCTCTGATTTGAGATGTGTGGGCTCTAATTTTGGGTATGATGACAA 3780
Q 3781 ATTTAACATTTGTGTTGTGCTCCCTACCGAGGAGCTCCCAAGTTTCTGATGAATAGA 3840
D 3781 ATTTAACATTTGTGTTGTGCTCCCTACCGAGGAGCTCCCAAGTTTCTGATGAATAGA 3840
Q 3841 CTGAAGAAATCAACAAGTGTCTATCTGCGCAGATTTAATGATATCTATTTCTTGCTT 3900
D 3841 CTGAAGAAATCAACAAGTGTCTATCTGCGCAGATTTAATGATATCTATTTCTTGCTT 3900
Q 3901 CTCCTCTCCCTGAGGACCTTAATTTATGTCCTCTCTAGGTAATTCCTCTTG 3960
D 3901 CTCCTCTCCCTGAGGACCTTAATTTATGTCCTCTCTCTAGGTAATTCCTCTTG 3960
Q 3961 ATTTGACTTGTGAGAGAGGTTGAGACAGTATGAGCAAGTTTCCAAAGTCAAAAT 4020
D 3961 ATTTGACTTGTGAGAGAGGTTGAGACAGTATGAGCAAGTTTCCAAAGTCAAAAT 4020
Q 4021 ACAAGTGTGAGAGTGTGGGGGAAAAATGCTTAATTTTCCCTACATGGGATACACA 4080
D 4021 ACAAGTGTGAGAGTGTGGGGGAAAAATGCTTAATTTTCCCTACATGGGATACACA 4080
Q 4081 CTGTGAATTAATCTTCAACTGAAGGCCCTGACAGTCTCTTAATAACATGATGTGTTT 4140
D 4081 CTGTGAATTAATCTTCAACTGAAGGCCCTGACAGTCTCTTAATAACATGATGTGTTT 4140
Q 4141 TTTCTTAACAAAGTTTAAGCTATGTTAATTAATTAATAAATGCTGTCTGTACT 4200
D 4141 TTTCTTAACAAAGTTTAAGCTATGTTAATTAATTAATAAATGCTGTCTGTACT 4200
Q 4201 TCAAGTTTGTATATGCCCAATTCATATGTTGTCTGTGTGTGAATTCATTAATCTTTGAT 4260
D 4201 TCAAGTTTGTATATGCCCAATTCATATGTTGTCTGTGTGTGAATTCATTAATCTTTGAT 4260
Q 4261 ACCATTTCTATGTGTAAATTTGTTGTCTTGTGAATATCTTAATAAGAGTTCAATGTA 4320
D 4261 ACCATTTCTATGTGTAAATTTGTTGTCTTGTGAATATCTTAATAAGAGTTCAATGTA 4320
Q 4321 AATTAACATTTGTGGCTGTAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAA 4372
D 4321 AATTAACATTTGTGGCTGTAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAA 4372

RESULT 2
US-10-144-194A-65
; Sequence 65, Application US/10144194A
; Publication NO. US20030215809A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Breast Cancer Genes

```
FILE REFERENCE: 3U 103 R1
CURRENT APPLICATION NUMBER: US/10/144,194A
CURRENT FILING DATE: 2002-06-12
NUMBER OF SEQ ID NOS: 114
SOFTWARE: PatentIn version 3.0
SEQ ID NO 65
LENGTH: 4372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (78) ..(1922)
US-10-144-194A-65

Query Match      100.0%; Score 4372; DB 17; Length 4372;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGTCTCGAACCCCGGACCTGCGCCCTTCCCAACCCCTCTCCGCTCGGCTGCGCGG 60
DB 1 CAGTCTCGAACCCCGGACCTGCGCCCTTCCCAACCCCTCTCCGCTCGGCTGCGCGG 60
QY 61 CGCTGCTCCGGAACCACTATGACCATGAGATCCGCGGTGTTCAAGGCGCGCGCCCTG 120
DB 61 CGCTGCTCCGGAACCACTATGACCATGAGATCCGCGGTGTTCAAGGCGCGCGCCCTG 120
QY 121 CGGCGCGCAATCTCTGAGACGCACTGACTGACGCGGCTGCGCGCTGCGCGCGCG 180
DB 121 CGGCGCGCAATCTCTGAGACGCACTGACTGACGCGGCTGCGCGCTGCGCGCGCG 180
QY 181 AGGACGAGCTGGGGGGGCGGAGCCCACTTCTCCCGCGACCCGTAAGCTCAAGAGC 240
DB 181 AGGACGAGCTGGGGGGGCGGAGCCCACTTCTCCCGCGACCCGTAAGCTCAAGAGC 240
QY 241 CGGCGCGCCCGCTGCTCTCTCCAGGCGGAGCCCGCGCTTCCCGCGCGCTGCG 300
DB 241 CGGCGCGCCCGCTGCTCTCTCCAGGCGGAGCCCGCGCTTCCCGCGCGCTGCG 300
QY 301 GCGGCAAGGCGCGGCGCTGTTACTCCCGCGCGGCGCGCGCGCGCGAGCAAGAG 360
DB 301 GCGGCAAGGCGCGGCGCTGTTACTCCCGCGCGGCGCGCGCGCGCGAGCAAGAG 360
QY 361 GCTGGGCGGCTTCTGAGCTTCCCTGCTCCGCGCGCGCGCGCGCGCGCGCGGATG 420
DB 361 GCTGGGCGGCTTCTGAGCTTCCCTGCTCCGCGCGCGCGCGCGCGCGCGCGGATG 420
QY 421 GGGGAGGCTGCGGAGCGGAGCGCGCTGACGCCCGCGCGCGCGCGCGCGCGGATG 480
DB 421 GGGGAGGCTGCGGAGCGGAGCGCGCTGACGCCCGCGCGCGCGCGCGCGCGGATG 480
QY 481 TCGCCATTGACGCGGCTCTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCTGCGG 540
DB 481 TCGCCATTGACGCGGCTCTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCTGCGG 540
QY 541 ACAAGGCTGGGGGGGCGCTCCCGCGCTGCCACCGCTTCGACCCGCGGCGGACCCCA 600
DB 541 ACAAGGCTGGGGGGGCGCTCCCGCGCTGCCACCGCTTCGACCCGCGGCGGACCCCA 600
QY 601 TACCTTGCCTGGGCGCGCACCCCTGCGCGCGCACCGCGCACCGCGCGGAGCCCTG 660
DB 601 TACCTTGCCTGGGCGCGCACCCCTGCGCGCGCACCGCGCACCGCGCGGAGCCCTG 660
QY 661 GCGAGGCGAGATGAGAGATATGAGAGAGCCCTCTCGGAGGTGTGTGCGGCTCGG 720
DB 661 GCGAGGCGAGATGAGAGATATGAGAGAGCCCTCTCGGAGGTGTGTGCGGCTCGG 720
QY 721 CCTCCAGTGAAGCGCGCTGCTCAACAGATCTTCTGCTGCAATTGGAATCTGAGAC 780
DB 721 CCTCCAGTGAAGCGCGCTGCTCAACAGATCTTCTGCTGCAATTGGAATCTGAGAC 780
QY 781 AGCAGCAGCAGCTGACAGCGCAAGAGAAAGAGATCGAGAGCTGGAAGTCAAGAGAG 840
DB 781 AGCAGCAGCAGCTGACAGCGCAAGAGAAAGAGATCGAGAGCTGGAAGTCAAGAGAG 840
```

```
QY 841 ACAAGCTCTTGTGCTGGATTGAAACGATATGGAAGAGCGGATGCAAGCTGTGTAAGAGAT 900
DB 841 ACAAGCTCTTGTGCTGGATTGAAACGATATGGAAGAGCGGATGCAAGCTGTGTAAGAGAT 900
QY 901 ACGAGAGAGAGAGCAAGCTGTGTTCAAGGCTATGAACTGAAGAGAGAGAGAGAGAG 960
DB 901 ACGAGAGAGAGAGCAAGCTGTGTTCAAGGCTATGAACTGAAGAGAGAGAGAGAGAG 960
QY 961 AGCTATCTGAAGAAATTAATCTGAAGTGCACCGGAGCTTTCCGAGACATCCAGACTC 1020
DB 961 AGCTATCTGAAGAAATTAATCTGAAGTGCACCGGAGCTTTCCGAGACATCCAGACTC 1020
QY 1021 TGCCTCCCAAGCCCTTCTATGTGGCGGAGTGGAAAGGGAACATTAAGAGAAATCCCAT 1080
DB 1021 TGCCTCCCAAGCCCTTCTATGTGGCGGAGTGGAAAGGGAACATTAAGAGAAATCCCAT 1080
QY 1081 TTGGAAGTACAGAGAGAGAGCTCCGTTAAAGCTGCTCTGAAATTTTCAAAAGTCA 1140
DB 1081 TTGGAAGTACAGAGAGAGAGCTCCGTTAAAGCTGCTCTGAAATTTTCAAAAGTCA 1140
QY 1141 AAACAAAACCTCTAAGCACTCTCTATTAAAGAGAACCTGTGTGTTCTTATCTGAAA 1200
DB 1141 AAACAAAACCTCTAAGCACTCTCTATTAAAGAGAACCTGTGTGTTCTTATCTGAAA 1200
QY 1201 CTGTTTGTAAAGTGAATTTGAGAGCCAGAGAAACCCAGAGAAAGCCCGGCTTCACTG 1260
DB 1201 CTGTTTGTAAAGTGAATTTGAGAGCCAGAGAAACCCAGAGAAAGCCCGGCTTCACTG 1260
QY 1261 ACACCCCAACAAAGCTCTCACTCCCAAAAGGAGCCAGACCCCATCCCAAGAGAGAG 1320
DB 1261 ACACCCCAACAAAGCTCTCACTCCCAAAAGGAGCCAGACCCCATCCCAAGAGAGAG 1320
QY 1321 CCTTCTCAAGTGAATGAGAAATTTGCGCTACTTTCACACAGAAATGTAATTTGTGTC 1380
DB 1321 CCTTCTCAAGTGAATGAGAAATTTGCGCTACTTTCACACAGAAATGTAATTTGTGTC 1380
QY 1381 GTTGGCAACAGCTCTCCCATCAAGCTTACATTTACGAGAAATCTCTCCAAAGAGAG 1440
DB 1381 GTTGGCAACAGCTCTCCCATCAAGCTTACATTTACGAGAAATCTCTCCAAAGAGAG 1440
QY 1441 AGACGTAGCAAGGTGTGATGCCATCAAGGTGTCAGAGAACTTCAAGTCTTGAGCTG 1500
DB 1441 AGACGTAGCAAGGTGTGATGCCATCAAGGTGTCAGAGAACTTCAAGTCTTGAGCTG 1500
QY 1501 TTCTTCTTGAAGGACCACTCAGTAGAGCTTCAAGGAGCCCAATCTTCAAGACTTT 1560
DB 1501 TTCTTCTTGAAGGACCACTCAGTAGAGCTTCAAGGAGCCCAATCTTCAAGACTTT 1560
QY 1561 TGGAGAACTGGATGACAGTGTGTTTTCGAGCGGCAATGCAAACTGAGCTGGATGAGA 1620
DB 1561 TGGAGAACTGGATGACAGTGTGTTTTCGAGCGGCAATGCAAACTGAGCTGGATGAGA 1620
QY 1621 AGAGAAAGAAAGATGGGATATTCAGAGGATGAGGGAACAAAGAAATTTTACAGCGACTG 1680
DB 1621 AGAGAAAGAAAGATGGGATATTCAGAGGATGAGGGAACAAAGAAATTTTACAGCGACTG 1680
QY 1681 AGCTCAGATGTAATAAAGAGAAAGAAATTCAGAGATCTGAGCTGAGTTAACCTCATTTT 1740
DB 1681 AGCTCAGATGTAATAAAGAGAAAGAAATTCAGAGATCTGAGCTGAGTTAACCTCATTTT 1740
QY 1741 TCCCTGAGCCAGATGATGTTGAAAGTTGATATTAACCCCTTCTTGTGTTGAGCAT 1800
DB 1741 TCCCTGAGCCAGATGATGTTGAAAGTTGATATTAACCCCTTCTTGTGTTGAGCAT 1800
QY 1801 TTGAGAGACATTAACAAATTAATCCCAAGAAATTTTGAAGTCACTCTGTTGAGAGAC 1860
DB 1801 TTGAGAGACATTAACAAATTAATCCCAAGAAATTTTGAAGTCACTCTGTTGAGAGAC 1860
QY 1861 GTAGCCGATGACAGATTTGAGATTCAGAGAGAACAAACCTCAACCGGAGCTGTAGAGAA 1920
DB 1861 GTAGCCGATGACAGATTTGAGATTCAGAGAGAACAAACCTCAACCGGAGCTGTAGAGAA 1920
```



```
|||||
Db 1501 TTCCTCTTGGAGGACCACTCAGTAGAGCTCTAAGGAGCCCAAACTCTTCAGACCTTT 1560
Qy 1561 TGGAGAACCTGGATGACAGTGTGTTTTCGAAGCGGCGATGCAAACTGAGACTGGATGAGA 1620
Db 1561 TGGAGAACCTGGATGACAGTGTGTTTTCGAAGCGGCGATGCAAACTGAGACTGGATGAGA 1620
Qy 1621 AGAGAAAGAAAGATGGGATATTCAAGAGATCAGGAAACAAGATTTTTCACAGCACTGC 1680
Db 1621 AGAGAAAGAAAGATGGGATATTCAAGAGATCAGGAAACAAGATTTTTCACAGCACTGC 1680
Qy 1681 AGCTCAGATGTATTAAGAAAGAAATTCAGGAATCTGAGCTGAGGTTACTCATTTT 1740
Db 1681 AGCTCAGATGTATTAAGAAAGAAATTCAGGAATCTGAGCTGAGGTTACTCATTTT 1740
Qy 1741 TCCCTAGGCGAGATGTTGAAGTTGATGATTCACCCCTCTTGCCCTGTGTAGCAT 1800
Db 1741 TCCCTAGGCGAGATGTTGAAGTTGATGATTCACCCCTCTTGCCCTGTGTAGCAT 1800
Qy 1801 TTGGAGACCATTAACCAAAATTAATCAACAGAAATTTTGTAGCTACCTGGTGGATGAGC 1860
Db 1801 TTGGAGACCATTAACCAAAATTAATCAACAGAAATTTTGTAGCTACCTGGTGGATGAGC 1860
Qy 1861 GTAGCCGATGACAGATTGGAGATCCAGAGAACCAACCTCAACCGGACGTGTAGGAAT 1920
Db 1861 GTAGCCGATGACAGATTGGAGATCCAGAGAACCAACCTCAACCGGACGTGTAGGAAT 1920
Qy 1921 AGCTGTGCTGGAGAAACCTGTCTTCAGATGTTGATGAGATGCCATTCGCGAGAGTGGC 1980
Db 1921 AGCTGTGCTGGAGAAACCTGTCTTCAGATGTTGATGAGATGCCATTCGCGAGAGTGGC 1980
Qy 1981 AGAGACCTGTATATGTGACCTTTGTCTCAGATATGTTATCACTGCGTGAATAACCTT 2040
Db 1981 AGAGACCTGTATATGTGACCTTTGTCTCAGATATGTTATCACTGCGTGAATAACCTT 2040
Qy 2041 TCATACCTCTTGAATGTTTCAATTAATCTGATTTTCAAAAACTCTTTCAATGGGCT 2100
Db 2041 TCATACCTCTTGAATGTTTCAATTAATCTGATTTTCAAAAACTCTTTCAATGGGCT 2100
Qy 2101 AATTGTGAGTTAGAGAGGATGGGATTTCTTTTCCCTTTTGGGAAATGGGCTCTC 2160
Db 2101 AATTGTGAGTTAGAGAGGATGGGATTTCTTTTCCCTTTTGGGAAATGGGCTCTC 2160
Qy 2161 AAGCTAAGCTATAGATGAGCAGATTCAGAAATTTCAAGGGGCTGTGTTCTATACATTTGC 2220
Db 2161 AAGCTAAGCTATAGATGAGCAGATTCAGAAATTTCAAGGGGCTGTGTTCTATACATTTGC 2220
Qy 2221 CTATGTTAAAGGGGTAAAGGGGCTCTCTTCAATTAAGACATGAGATGAGAGACCCCT 2280
Db 2221 CTATGTTAAAGGGGTAAAGGGGCTCTCTTCAATTAAGACATGAGATGAGAGACCCCT 2280
Qy 2281 TCCCTTAGAGCTGTGCTGATGAGCACTTTCTCAACCTGTGACACCTCTCTTATAGTGG 2340
Db 2281 TCCCTTAGAGCTGTGCTGATGAGCACTTTCTCAACCTGTGACACCTCTCTTATAGTGG 2340
Qy 2341 GTATAGGATTTTAAACCTTAAATAAACAACAACCTGACCACTGAGCTTTTGGACAG 2400
Db 2341 GTATAGGATTTTAAACCTTAAATAAACAACAACCTGACCACTGAGCTTTTGGACAG 2400
Qy 2401 AAGAGAAATGACAAGTGAAGCAGATGAGCAAGCATTTTCAAGAGTGAAGAAAGACATCG 2460
Db 2401 AAGAGAAATGACAAGTGAAGCAGATGAGCAAGCATTTTCAAGAGTGAAGAAAGACATCG 2460
Qy 2461 GAGAGTTGTAGATTAATCTGTCTGAAAAAGATGTTGCTCATTTGAACTATCTGTGTATC 2520
Db 2461 GAGAGTTGTAGATTAATCTGTCTGAAAAAGATGTTGCTCATTTGAACTATCTGTGTATC 2520
Qy 2521 AGTCATGTGGGAAGGAGTGTGTGGCTGTGATTAATTTTTCAGTTAATGATTAACAATTC 2580
Db 2521 AGTCATGTGGGAAGGAGTGTGTGGCTGTGATTAATTTTTCAGTTAATGATTAACAATTC 2580
Qy 2581 TTTACTGTCAAAAACCAAAATCTTTTGGAAAAGAAAGTGGGATGTTAGTTTCAGAAC 2640
Db 2581 TTTACTGTCAAAAACCAAAATCTTTTGGAAAAGAAAGTGGGATGTTAGTTTCAGAAC 2640
|||||
Db 2641 AGTTACAGCTGTATAACCAAAAGCACTTAATGATTTGGATGGATGCCAAAACCTGTATAAA 2700
Qy 2641 AGTTACAGCTGTATAACCAAAAGCACTTAATGATTTGGATGGATGCCAAAACCTGTATAAA 2700
Db 2701 TGTCTGTATACATCACTTCTCAAGTATTTCTTCAATTTGGGCTTCACTCTTTTACAGA 2760
Qy 2701 TGTCTGTATACATCACTTCTCAAGTATTTCTTCAATTTGGGCTTCACTCTTTTACAGA 2760
Db 2761 ACTCTGTGTGGGATAGAGACTTGAAGAGGATAGAGGAGATGAGGAAATAGGTGCT 2820
Qy 2761 ACTCTGTGTGGGATAGAGACTTGAAGAGGATAGAGGAGATGAGGAAATAGGTGCT 2820
Db 2821 TCCCTTGGCTGCAAAATGTCTACATCTTTGAAACAAACAGATGTACTAATGACCTTCC 2880
Qy 2821 TCCCTTGGCTGCAAAATGTCTACATCTTTGAAACAAACAGATGTACTAATGACCTTCC 2880
Db 2881 ATTCACTTTGTAAAAATTTGTATGTGTACATCTTGGCTCTTCCCTCCGTTTTG 2940
Qy 2881 ATTCACTTTGTAAAAATTTGTATGTGTACATCTTGGCTCTTCCCTCCGTTTTG 2940
Db 2941 ATTCACTTTGTAAAAATTTGTATGTGTACATCTTGGCTCTTCCCTCCGTTTTG 2940
Qy 2941 TTTAAATATCAGATAGCACTCCAGGCCACTTTGGTCTCAGTGAAGATCCCTATTAAC 3000
Db 2941 TTTAAATATCAGATAGCACTCCAGGCCACTTTGGTCTCAGTGAAGATCCCTATTAAC 3000
Qy 3001 TATCTGAAGAAATATAGAGCAAGCACTTGTGTCTCAATATATATAGAAATGCTTTCT 3060
Db 3001 TATCTGAAGAAATATAGAGCAAGCACTTGTGTCTCAATATATATAGAAATGCTTTCT 3060
Qy 3061 TTAGTCTTCAGACATAATGTGTGAAAAACAATGAGGGGTCTAATCTCTAGAGATGAGG 3120
Db 3061 TTAGTCTTCAGACATAATGTGTGAAAAACAATGAGGGGTCTAATCTCTAGAGATGAGG 3120
Qy 3121 CTATTATCTTTAAAGGATATGTCCCAGATTAATTAAGCACTTTTGAAGAGAACCAAG 3180
Db 3121 CTATTATCTTTAAAGGATATGTCCCAGATTAATTAAGCACTTTTGAAGAGAACCAAG 3180
Qy 3181 GTATGTAGGGGTGTGGCTGGGCCATCAGTGGAGCAAGAAAGAGATGGGATACCATTTG 3240
Db 3181 GTATGTAGGGGTGTGGCTGGGCCATCAGTGGAGCAAGAAAGAGATGGGATACCATTTG 3240
Qy 3241 TGGGAAGAGAAAGAAAGTTCTTCAGGGGCTCCCACTGTCTAAAGTTTGTGAGATGT 3300
Db 3241 TGGGAAGAGAAAGAAAGTTCTTCAGGGGCTCCCACTGTCTAAAGTTTGTGAGATGT 3300
Qy 3301 GATCTGTGCTCTCGAATTTGACTTTTAAAGAAATTAATTTGCGACACATGTATATTC 3360
Db 3301 GATCTGTGCTCTCGAATTTGACTTTTAAAGAAATTAATTTGCGACACATGTATATTC 3360
Qy 3361 TTGGAGATCTGCTGCTCTTATTTCTCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3420
Db 3361 TTGGAGATCTGCTGCTCTTATTTCTCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3420
Qy 3421 TGGGTTTTCATTTGTAACTCATCTGCTTAAGAGAGTGGGCTCTCTTAAGAGAACTGC 3480
Db 3421 TGGGTTTTCATTTGTAACTCATCTGCTTAAGAGAGTGGGCTCTCTTAAGAGAACTGC 3480
Qy 3481 TGTAAACTTATTTGACGAAGAGATGTAGAGAAATATAGAACTTAATTTCCACTGAGGCTC 3540
Db 3481 TGTAAACTTATTTGACGAAGAGATGTAGAGAAATATAGAACTTAATTTCCACTGAGGCTC 3540
Qy 3541 TCAATTCACACCTTAAGAGAGATTTCTTAAGAAACTGGGCCAGATTTCTTGTGTCTC 3600
Db 3541 TCAATTCACACCTTAAGAGAGATTTCTTAAGAAACTGGGCCAGATTTCTTGTGTCTC 3600
Qy 3601 CATCATTTTATGTGCAAGGCTGTTCAGTTTCTTACTCTTACATATGTATATTTCTTC 3660
Db 3601 CATCATTTTATGTGCAAGGCTGTTCAGTTTCTTACTCTTACATATGTATATTTCTTC 3660
Qy 3661 GTACGCTGTCAAAAAGAAAGAACCAATCAAGTGTCTTTGACATTTGTTGATTC 3720
Db 3661 GTACGCTGTCAAAAAGAAAGAACCAATCAAGTGTCTTTGACATTTGTTGATTC 3720
|||||
```

QY	3721	CTGAGTTTCTTCCTGAATTTCAAGCATGTGGCGGTTCCAAATTTGGGGTAAGATTACAA	3780
Db	3721	CTCAGTTTTCTTCTTGATTTTCAGCATGTGTGGCGGTTCTCAAATTTGGGGTAAGATTACAA	3780
QY	3781	ATTTAACCAATGTGTGTGTGGCCCTACCAGAAGGAGCTCCCCAGTTTCTGACTTGAAGTAGA	3840
Db	3781	ATTTAACCAATGTGTGTGTGGCCCTACCAGAAGGAGCTCCCCAGTTTCTGACTTGAAGTAGA	3840
QY	3841	CTGAGAAAGAAATCCACGAGGTGCTATCTGGCCAGATTAAAGTAAGTTCTATTTCTTGCTT	3900
Db	3841	CTGAGAAAGAAATCCACGAGGTGCTATCTGGCCAGATTAAAGTAAGTTCTATTTCTTGCTT	3900
QY	3901	CTGCCCTCTCCCTGAGAGCTCTTATTTTAAATGTGCCCTCTCPAAGTTAATTCCTTTG	3960
Db	3901	CTGCCCTCTCCCTGAGAGCTCTTATTTTAAATGTGCCCTCTCPAAGTTAATTCCTTTG	3960
QY	3961	ATTGTAAGTTGTGAGAGAGAGTTGAGCAGTAGATTAGCAAAAGTCCAAAGTGCAAAATT	4020
Db	3961	ATTGTAAGTTGTGAGAGAGAGTTGAGCAGTAGATTAGCAAAAGTCCAAAGTGCAAAATT	4020
QY	4021	ACAAGTGTGTAGAGTGTGGGGGAAAAATTAGTCTTATTTTTCCTACATGGGAATACAA	4080
Db	4021	ACAAGTGTGTAGAGTGTGGGGGAAAAATTAGTCTTATTTTTCCTACATGGGAATACAA	4080
QY	4081	CTGGAATTTCAATCTTCAACCTGAAGGCGCCGTCAGTCTCCATAAAACATAGTTGTTGTT	4140
Db	4081	CTGTGAATTTCAATTTTCAACCTGAAGGCGCCGTCAGTCTCTTAATAACATAGTTGTTGTT	4140
QY	4141	TTCCTTTACAAAAGTTTAAAGCTAGTGTAAATTAATTAATAAAAAAAAAATGCTGTGTCTACT	4200
Db	4141	TTCCTTTACAAAAGTTTAAAGCTAGTGTAAATTAATTAATAATAAAAAAAAAATGCTGTGTCTACT	4200
QY	4201	TCAGCTTGTGTTTATGCCCATTTCAATGTGTGTCTGTGTGTGAATCATPACTTTTGAT	4260
Db	4201	TCAGCTTGTGTTTATGCCCATTTCAATGTGTGTGTGTGTGAATCATPACTTTTGAT	4260
QY	4261	ACCATTTCTGATGTGTAATAATTTGGTGTCTGTGAATTCCTTATPAAGAGTCAATGTGA	4320
Db	4261	ACCATTTCTGATGTGTAATAATTTGGTGTCTGTGAATTCCTTATPAAGAGTCAATGTGA	4320
QY	4321	AATTAACATATTTGTGGCTGTGTAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	4372
Db	4321	AATTAACATATTTGTGGCTGTGTAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	4372

RESULT 4
US-10-641-643-167
; Sequence 167, Application US/10641643
; Publication No. US20040077003A1
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
Susan G. Stuart
Jeffrey J. Seilhaeuer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
GENE EXPRESSION
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/641,643
FILING DATE: 14-Aug-2003
CLASSIFICATION: <Unknown>

Query	Match	34.5%	Score 1508.4	DB 17	Length 1545
Db	1	GGGGGAGAGTGTGGAATAAGTGTCTTGGCTGGCAATGCTACATCTTGAACAA	2855		
Qy	2796	GGGGGAGAGTGTGGAATAAGTGTCTTGGCTGGCAATGCTACATCTTGAACAA	2855		
Db	1	GGGGGAGAGTGTGGAATAAGTGTCTTGGCTGGCAATGCTACATCTTGAACAA	60		
Qy	2856	ACAGATGTACCTAATGAGCTTCTTCACATCTTGTAAAAATAATTGTATGTACCAT	2915		
Db	61	ACAGATGTACCTAATGAGCTTCTTCACATCTTGTAAAAATAATTGTATGTACCAT	120		
Qy	2916	CTTGATCTCTCCCTCCGCTTTTGTAAAAATACAGATAGACATCCGAGCCACTTG	2975		
Db	121	CTTGATCTCTCCCTCCGCTTTTGTAAAAATACAGATAGACATCCGAGCCACTTG	180		
Qy	2976	GTCCTAGTGTAAAGTCCCTATTAACTATCTGAAGAAAAATAGACCAGACCTTGTC	3035		
Db	181	GTCCTAGTGTAAAGTCCCTATTAACTATCTGAAGAAAAATAGACCAGACCTTGTC	240		
Qy	3036	TCAATATATAGAAATGTGCTTCTTTAATGTCTTACAGATATTTGTGAAAAACAATAG	3095		
Db	241	TCAATATATAGAAATGTGCTTCTTTAATGTCTTACAGATATTTGTGAAAAACAATAG	300		
Qy	3096	GGTCTATCTCTAGAAAGTAGAGGGCTTTATCTTAAAGAAATATGTCCCGCATTTAT	3155		
Db	301	GGTCTATCTCTAGAAAGTAGAGGGCTTTATCTTAAAGAAATATGTCCCGCATTTAT	360		
Qy	3156	TAGCACTTTTGAAGAGAACCAAGATATGAGGGTGTGGCTGGCCATCACTGAGC	3215		
Db	361	TAGCACTTTTGAAGAGAACCAAGATATGAGGGTGTGGCTGGCCATCACTGAGC	420		
Qy	3216	ACGAAGAGAAATGGGATACCATTTGGGAAAGAAATAAATGTCCTCAGGGGCTCCCA	3275		
Db	421	ACGAAGAGAAATGGGATACCATTTGGGAAAGAAATAAATGTCCTCAGGGGCTCCCA	480		
Qy	3276	CTGCTAAAGTTTTTGTGAGATGTGATCTGTGCTTCTGAGATTTGACTTTAAAGAAAT	3335		
Db	481	CTGCTAAAGTTTTTGTGAGATGTGATCTGTGCTTCTGAGATTTGACTTTAAAGAAAT	540		
Qy	3336	TATTTGGGACACATGTAGTATTTCTTGATGATCTTGCTCTTAAATTTCTCTTTGT	3395		
Db	541	TATTTGGGACACATGTAGTATTTCTTGATGATCTTGCTCTTAAATTTCTCTTTGT	600		
Qy	3396	GTCGTGTGTGTGTGTGTGTGGCTATAGGGTTTTCATTTGAACTCCCATCTGCTTAGAGA	3455		
Db	601	NN	660		
Qy	3456	GTCGGCTCTCTATTAAGGAACCTGCTGTAAATTCATTTGAGCAAGAGTATAGAGAAA	3515		
Db	661	GTCGGCTCTCTATTAAGGAACCTGCTGTAAATTCATTTGAGCAAGAGTATAGAGAAA	720		

QY	3516	TAGAGACTTAATTCACATAGGGGGCTCATCTCAACACCTTAAGAGAGATTCTAGAAAA	3575
Db	721	TAGAGACTTAATTCACATAGGGGGCTCATCTCAACACCTTAAGAGAGATTCTAGAAAA	780
QY	3576	ACTGGGCGAGATTTTCTTTGTTCTTCACATCAATTTTAATGAGGACGGCTGTCAAGTTTCTT	3635
Db	781	ACTGGGCGAGATTTTCTTTGTTCTTCACATCAATTTTAATGAGGACGGCTGTCAAGTTTCTT	840
QY	3636	ACTCTTAACCTATGATATATTTCTTCCTGATACGCTGTCACAAAAGAAAAAGACCACATCAGT	3695
Db	841	ACTCTTAACCTATGATATATTTCTTCCTGATACGCTGTCACAAAAGAAAAAGACCACATCAGT	900
QY	3696	GTCCTCTGAACCTTGTCTTTGATCCCTCAGTTTCTTCTTGATTTTCAGCATGTGTCCGGTT	3755
Db	901	GTCCTCTGAACCTTGTCTTTGATCCCTCAGTTTCTTCTTGATTTTCAGCATGTGTCCGGTT	960
QY	3756	CCTAATTTTGGGATGAGTTAGTACAAATTTAACATGTGTGTTGTGCCCTTACCCAGGGAC	3815
Db	961	CCTAATTTTGGGATGAGTTAGTACAAATTTAACATGTGTGTTGTGCCCTTACCCAGGGAC	1020
QY	3816	TCCCCAGTTTCTGACTTGAATGATACGTAGAAAGATCCACAGAGTGTCTATCTGGCCAGAT	3875
Db	1021	TCCCCAGTTTCTGACTTGAATGATACGTAGAAAGATCCACAGAGTGTCTATCTGGCCAGAT	1080
QY	3876	TTAAGTAGAATCTATTTCTTGTTCTCCCTCTCCCTCAGGACCCCTAATTTTAATTTATGCC	3935
Db	1081	TTAAGTAGAATCTATTTCTTGTTCTCCCTCTCCCTCAGGACCCCTAATTTTAATTTATGCC	1140
QY	3936	CCTCTTCTAGGTAAATTCCTCTTGATTTGACTTTGTTGAGAGAGAGTTGAGACATAGA	3995
Db	1141	CCTCTTCTAGGTAAATTCCTCTTGATTTGACTTTGTTGAGAGAGAGTTGAGACATAGA	1200
QY	3996	TTACCAAAAGTTCCCAAGTGCACAAATTTACAGTGTGTTAAGTGTGGGGGAAAAATTAGCTTT	4055
Db	1201	TTACCAAAAGTTCCCAAGTGCACAAATTTACAGTGTGTTAAGTGTGGGGGAAAAATTAGCTTT	1260
QY	4056	ATTTTTCCTCCATCAGGAGATACACACTGTGAATTCATCTTCAACTGAAAGCCCTGCAGT	4115
Db	1261	ATTTTTCCTCCATCAGGAGATACACACTGTGAATTCATCTTCAACTGAAAGCCCTGCAGT	1320
QY	4116	TCTCCTAAACATATGTTGTTGTTTTCTTTTAAACAAAGTTTAAGCTAGTGTAAATAAT	4175
Db	1321	TCTCCTAAACATATGTTGTTGTTTTCTTTTAAACAAAGTTTAAGCTAGTGTAAATAAT	1380
QY	4176	AAAAAAAATTCCTGTCTGTCTACTCTCAGCTTTGTTTATGCCCCATTTCAATATGTGTCTC	4235
Db	1381	AAAAAAAATTCCTGTGTCTGTCTACTCTCAGCTTTGTTTATGCCCCATTTCAATATGTGTCTC	1440
QY	4236	TGTGTGTGAATTCATTAACCTTTTGATACCATTTTCGATGTGTAAAAATGGTGTCTTGATA	4295
Db	1441	TGTGTGTGAATTCATTAACCTTTTGATACCATTTTCGATGTGTAAAAATGGTGTCTTGATA	1500
QY	4296	ATATCTTATATAAAGATTCAATTGTAAATTAACATAATGTGGCTGTT 4340	
Db	1501	ATATCTTATATAAAGATTCAATTGTAAATTAACATAATGTGGCTGTT 1545	
RESULT 5			
US-09-397-945-79			
Sequence 79, Application US/09397945			
Publication No. US20030065139A1			
GENERAL INFORMATION:			
APPLICANT: Human Genome Sciences, Inc. et al.			
TITLE OR INVENTION: 95 Human secreted proteins			
FILE REFERENCE: P2027P1			
CURRENT APPLICATION NUMBER: US/09/397,945			
PRIOR FILING DATE: 1999-09-17			
PRIOR APPLICATION NUMBER: PCT/US99/05804			
PRIOR FILING DATE: 1999-03-18			
PRIOR APPLICATION NUMBER: 60/078,566			
PRIOR FILING DATE: 1998-03-19			
PRIOR APPLICATION NUMBER: 60/078,576			

```

? PRIOR FILING DATE: 1998-03-19
? PRIOR APPLICATION NUMBER: 60/078,573
? PRIOR FILING DATE: 1998-03-19
? PRIOR APPLICATION NUMBER: 60/078,574
? PRIOR FILING DATE: 1998-03-19
? PRIOR APPLICATION NUMBER: 60/078,579
? PRIOR FILING DATE: 1998-03-19
? PRIOR APPLICATION NUMBER: 60/080,314
? PRIOR FILING DATE: 1998-04-01
? PRIOR APPLICATION NUMBER: 60/080,312
? PRIOR FILING DATE: 1998-04-01
? PRIOR APPLICATION NUMBER: 60/078,578
? PRIOR FILING DATE: 1998-03-19
? PRIOR APPLICATION NUMBER: 60/078,561
? PRIOR FILING DATE: 1998-03-19
? PRIOR APPLICATION NUMBER: 60/078,577
? PRIOR FILING DATE: 1998-03-19
? PRIOR APPLICATION NUMBER: 60/078,563
? PRIOR FILING DATE: 1998-03-19
? PRIOR APPLICATION NUMBER: 60/080,313
? PRIOR FILING DATE: 1998-04-01
? NUMBER OF SEQ ID NOS: 470
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 79
? LENGTH: 1509
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-397-945--79

```

Query Match	33.8%	Score 1478;	DB 10;	Length 1503;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches 1500; Conservative	0;	Mismatches 0;	Indels 2;	Gaps 2;

QY	2859	GATGTACCTTAATGAGCTTCCTCCATCACTTGTGTAATAATATTTGATGTGTACATCTT	291.8
Db	9	GATGTACTTAATGAGCTTCCTCCATCACTTGTGTAATAATATTTGATGTGTACATCTT	68
QY	2919	GGTCCTCTCCCTCCGCGTGTGTTAAATAATCAGGATAGCACTCCAGGCACTTTTGCTC	297.7
Db	69	GGTCCTCTCCCTCCGCGTGTGTTAAATAATCAGGATAGCACTCCAGGCACTTTTGCTC	128
QY	2979	TCACTGTAAAGATCCCTATTAACTATCTGAAAAGAAAATAGAGCCAAAGCTCTGTCTCA	303.8
Db	129	TCACTGTAAAGATCCCTATTAACTATCTGAAAAGAAAATAGAGCCAAAGCTCTGTCTCA	188
QY	3039	AATATATAGGAATGGCCCTTCTTAGCTTCAGAGCTATTGTGTGAAAACAAGTAGGGGT	309.8
Db	189	AATATATAGGAATGGCCCTTCTTAGCTTCAGAGCTATTGTGTGAAAACAAGTAGGGGT	248
QY	3099	CTAATCTCTAGAAAGTAGGGGCTTTATCCTTAAAGAAATATGTCCCCAGTTATTAG	315.8
Db	249	CTAATCTCTAGAAAGTAGGGGC-TTTATCCTTAAAGAAATATGTCCCCAGTTATTAG	307
QY	3159	CACTTTTAGAGGAGAACCAAGGATATGTA-GGGTGTGTGCTGGCCCATCAGTGAAGAC	321.7
Db	308	CACTTTTAGAGGAGAACCAAGGATATGTAAGGGGTGTGTGCTGGCCCATCAGTGAAGAC	367
QY	3218	GAAAGAGGAATGGGATACCATTTGTGGGAAAGAAAGAAAGTTCTCCAGGGGCTCCCACT	327.7
Db	368	GAAAGAGGAATGGGATACCATTTGTGGGAAAGAAAGAAAGTTCTCCAGGGGCTCCCACT	427
QY	3278	GCTAAAGTTTTTGTGAGATGTTATCATCTGTGCTCTCGAGTTTGACTTTTAAAGGAATTA	333.7
Db	428	GCTAAAGTTTTTGTGAGATGTTATCATCTGTGCTCTCGAGTTTGACTTTTAAAGGAATTA	487
QY	3338	TTCTGGAGAGCAATGATAGTATCTTGGAGTAATCTTGCTCTTATTTTCTCTTTTGCT	339.7
Db	488	TTCTGGAGAGCAATGATAGTATCTTGGAGTAATCTTGCTCTTATTTTCTCTTTTGCT	547
QY	3398	GTTGTGTGTGTGTGTGTGTGTGCTAATGGGTTTTCAATTTGTAACTCCATCTGCTTAGAGAGT	345.5
Db	548	GTTGTGTGTGTGTGTGTGTGTGCTAATGGGTTTTCAATTTGTAACTCCATCTGCTTAGAGAGT	607

D	b		548	GTCGTGCTGTGTCGTGGCTAATGGATTTCATTGTAACTCCATCTGTTAGGAAGT	607
Q	y		3458	GGGCTCTCTTAAGAAGAACCTGCTGTAAACCTTCATTTGACAGCAAGATGTAGAAGAAATA	3517
D	b		608	GGGCTCTCTTAAGAAGAACCTGCTGTAAACCTTCATTTGACAGCAAGATGTAGAAGAAATA	667
Q	y		3518	GGACTTAATTCACTAGAGGGCTCATCTCACACTTAAGAGGAGATTTCTAGAAAAC	3577
D	b		668	GGACTTAATTCACTAGAGGGCTCATCTCACACTTAAGAGGAGATTTCTAGAAAAC	727
Q	y		3578	TGGGCCAGATTTTCTTGTTCTCCATCATATTTAAATGTGGCAGGCTGTTCAAGTTTCTTAC	3637
D	b		728	TGGGCCAGATTTTCTTGTTCTCCATCATATTTAAATGTGGCAGGCTGTTCAAGTTTCTTAC	787
Q	y		3638	TCTTAACCTATGTGATATTTCTTGCGTAACGTGTCCAAAAAGAAAAGCCATACAGT	3697
D	b		788	TCTTAACCTATGTGATATTTCTTGCGTAACGTGTCCAAAAAGAAAAGCCATACAGT	847
Q	y		3698	CTCTTGACTTGTCTTGATACCCTCAGATTCCTTGAATTCAGCATGTGTGGGTTCC	3758
D	b		848	CTCTTGACTTGTCTTGATACCCTCAGATTCCTTGAATTCAGCATGTGTGGGTTCC	907
Q	y		3758	TAAATTTGGGTATGAGTTAGCAAAATTTAAACCATGTGTGTTGTGCCCTAACCGAGGAGATC	3817
D	b		908	TAAATTTGGGTATGAGTTAGCAAAATTTAAACCATGTGTGTTGTGCCCTAACCGAGGAGATC	967
Q	y		3818	CCCAGTTTCTGACTTGAAGTATGACTGAGAAAGATCCACAGAGTGTATCTGGCCAGATTT	3877
D	b		968	CCCAGTTTCTGACTTGAAGTATGACTGAGAAAGATCCACAGAGTGTATCTGGCCAGATTT	1027
Q	y		3878	AAGTAGATTTCAATTTCCCTGGTCTGCCCTCCCTGAGAGACCTCTAATTTATGTCCCC	3937
D	b		1028	AAGTAGATTTCAATTTCCCTGGTCTGCCCTCCCTGAGAGACCTCTAATTTATGTCCCC	1087
Q	y		3938	TCTTCTAGGTTAATTTCTCTTTGATTTTGAATTTGACCTTTGTTGAGAAAGAGTTGACAGTAGATT	3997
D	b		1088	TCTTCTAGGTTAATTTCTCTTTGATTTTGAATTTGACCTTTGTTGAGAAAGAGATTGACAGTAGATT	1147
Q	y		3998	AGCAAAGTCCAAAGTCGAAATTTACAGTGTGTAGAGTGTGGGGGAAAAATTAAGCTTAT	4057
D	b		1148	AGCAAAGTCCAAAGTCCAAATTTACAGTGTGTAGAGTGTGGGGGAAAAATTAAGCTTAT	1207
Q	y		4058	TTTTTCCCTACATGGGATPACAACATGTGAATTCATCTTCAACTGAAGAGCCCTGCAGTTC	4117
D	b		1208	TTTTTCCCTACATGGGATPACAACATGTGAATTCATCTTCAACTGAAGAGCCCTGCAGTTC	1267
Q	y		4118	TCCGTAACACATAGTTGTTGTTTTCTTTTAACAAAAGTTAAAGCTAGTGTAAATAATTA	4177
D	b		1268	TCCGTAACACATAGTTGTTGTTTTCTTTTAACAAAAGTTAAAGCTAGTGTAAATAATTA	1327
Q	y		4178	AAAAAATTTGCTGTCTGTCTACCTTCAGCTTGTGTTTATGCCCATTTCATATTTGTGTCTG	4237
D	b		1328	AAAAAATTTGCTGTCTGTCTACCTTCAGCTTGTGTTTATGCCCATTTCATATTTGTGTCTG	1387
Q	y		4238	TGTTGTATTAATPAPAACTTTTGATPACAATTCCTGATGTGTAAATTTGATGCTTGTAAT	4297
D	b		1388	TGTTGTATTAATPAPAACTTTTGATPACAATTCCTGATGTGTAAATTTGATGCTTGTAAT	1447
Q	y		4298	ATCTTAATAAGAGTTCAAATTTGTAATAATAAATAATTTGTGGCTGTGTAAAAA	4357
D	b		1448	ATCTTAATAAGAGTTCAAATTTGTAATAATAAATAATTTGTGGCTGTGTAAAAA	1507
Q	y		4358	AA 4359	
D	b		1508	AA 1509	

RESULT 7
US-09-925-302-212
; Sequence 212, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.

```

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925.302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 212
; LENGTH: 1529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-302-212

Query Match          19.1%; Score 837; DB 9; Length 1529;
Best Local Similarity 100.0%; Pred. NO. 1.2e-203;
Matches 837; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

OY	617	GCACACCCCTCGCGCCACCGCCACCGCCGGAGACCTTGGCGGCGCAGCGAGGGCAGATGGAA	676
Db	2	GCACACCCCTCGCGCCACCGCCACCGCCGGAGACCTTGGCGGCGCAGCGAGGGCAGATGGAA	61
OY	677	GAGTATGAGGAAGACCCCTCTCGGGGGGTGTGGCGGCTCGGGAGCTTCAGTCAGGCCGC	736
Db	62	GAGTATGAGGAAGACCCCTCTCGGGGGGTGTGGCGGCTCGGGAGCTTCAGTCAGGCCGC	121
OY	737	CTGCTTCAAAAGATCTTCTGTGTGTGCMAATTGGACTTCATTCGACACGACGACGACACT	796
Db	122	CTGCTTCAAAAGATCTTCTGTGTGTGCMAATTGGACTTCATTCGACACGACGACGACACT	181
OY	797	GCAGGCCCAAGAAAGAGATCGAGAGCTGAAGTCAGAGAGAGACACGCTCTCTTGCTCG	856
Db	182	GCAGGCCCAAGAAAGAGATCGAGAGCTGAAGTCAGAGAGAGACACGCTCTCTTGCTCG	241
OY	857	GATTGAACGTATGAAAAGCGGATGCGAGCTGTATAAGAGATPACGAAAGAAAGGCA	916
Db	242	GATTGAACGTATGAAAAGCGGATGCGAGCTGTATAAGAGATPACGAAAGAAAGGCA	301
OY	917	CAACCTGTTTCAGGGCTTATGAACTTGAAAGAGAGAGAAACAGAGCTTATCTGAGAAAT	976
Db	302	CAACCTGTTTCAGGGCTTATGAACTTGAAAGAGAGAGAAACAGAGCTTATCTGAGAAAT	361
OY	977	TAACTGGAGTGCAGCCGGAGCTTTCGAGACATCCGAGACTGCTCCCAAGCCCTT	1038
Db	362	TAACTGGAGTGCAGCCGGAGCTTTCGAGACATCCGAGACTGCTCCCAAGCCCTT	421
OY	1037	CTCATGTGTGGCGGAGTGGAAAGGGAATPATAAAGGAATCCCATTTGGAGTACGAAAG	1096
Db	422	CTCATGTGTGGCGGAGTGGAAAGGGAATPATAAAGGAATCCCATTTGGAGTACGAAAG	481
OY	1097	AAAGATCTCTGTAAAAAGCTGTGCTCTCTGAATTTTCAAAAGTCAAAACAAAAATCTCTTA	1155
Db	482	AAAGATCTCTGTAAAAAGCTGTGCTCTCTGAATTTTCAAAAGTCAAAACAAAAATCTCTTA	541
OY	1157	GCACCTCTCTATTTAAAGAGAAACCTGTGGTCTCTTATCTGAAACTGTTTGTAAACGTGA	1216
Db	542	GCACCTCTCTATTTAAAGAGAAACCTGTGGTCTCTTATCTGAAACTGTTTGTAAACGTGA	601
OY	1217	ATTGAGGAGCCAGAAACCCCAAGAAAAGCCCGGTCTTCAGTGGACACCCACCAAGACT	1276
Db	602	ATTGAGGAGCCAGAAACCCCAAGAAAAGCCCGGTCTTCAGTGGACACCCACCAAGACT	661
OY	1277	CTCCACCTCCCAAAAGGAGCCGACACCCATCCCAAGAGAAAGCTTCTCAAGTAGAAT	1338
Db	662	CTCCACCTCCCAAAAGGAGCCGACACCCATCCCAAGAGAAAGCTTCTCAAGTAGAAT	721
OY	1337	AGAGATTTGGCGGTACTTTTCACACACAGAAATGTATTTGTGCTTGGACCAAGCCCTCC	1398
Db	722	AGAGATTTGGCGGTACTTTTCACACACAGAAATGTATTTGTGCTTGGACCAAGCCCTCC	781
OY	1397	CCCATTCACCGTTACCATTTACGGGAATCTCTCCAAAGAGAGAGAGACTGTAGCAG	1453


```

Db      61 AAGAAAGCAAGAGCTGTTTCAGGGCTTATGAACTGAGAGAGAGAGAAACAGACTGT 120
Qy      967 CTGAGAAAATTAATCTGAGAGTGCACGCGAGCTTCCGAGACATCCAGACTCTGCTC 1026
Db      121 CTGAGAAAATTAATCTGAGAGTGCACGCGAGCTTCCGAGACATCCAGACTCTGCTC 180
Qy      1027 CCAAGCCCTTCTCATGTGGCGGAGT-GGAAAGGAGCATTAAGAAATGCCATTGGA 1085
Db      181 CCAAGCCCTTCTCATGTGGCGGAGTGGGAAAGGAGCATTAAGAAATGCCATTGGA 240
Qy      1086 AGTACGAAAGAAAGAGCTCTGTTAAAGAGTGGCTCTGGAATTTCAAAAGTCAAAACA 1145
Db      241 AGTACGAAAGAAAGAGCTCTGTTAAAGAGTGGCTCTGGAATTTCAAAAGTCAAAACA 300
Qy      1146 AAAACCTCTAGACCTCTCTATTAAGAGAACCCGTGGTCTTATCTGAAGCTGT 1205
Db      301 AAAACCTCTAGACCTCTCTATTAAGAGAACCCGTGGTCTTATCTGAAGCTGT 360
Qy      1206 TGTAAACGTGAATTGAGAGCCCAAGAAACCCCAAGAAAGCCGCTTCAAGTGA 1265
Db      361 TGTAAACGTGAATTGAGAGCCCAAGAAACCCCAAGAAAGCCGCTTCAAGTGA 420
Qy      1266 CCAAGCAAGCTCTCCACTCCCAAAAGGAGCCAGACCCATCCAAAGAGAAAGCTTC 1325
Db      421 CCAAGCAAGCTCTCCACTCCCAAAAGGAGCCAGACCCATCCAAAGAGAAAGCTTC 480
Qy      1326 TCAAGTGAATGAAGATTGGCCGTACTTCCACCAAGAAATGATTTGTCGTTG 1385
Db      481 TCAAGTGAATGAAGATTGGCCGTACTTCCACCAAGAAATGATTTGTCGTTG 540
Qy      1386 CACCAAGCTCTCCACTCCCAAGGATTAAGGAAATCTCTCAAGAGAGAGAGACT 1445
Db      541 CACCAAGCTCTCCACTCCCAAGGATTAAGGAAATCTCTCAAGAGAGAGAGACT 600
Qy      1446 GTAGCAAG 1453
Db      601 GTAGCAAG 608

RESULT 10
US-10-242-535A-51339
; Sequence 51339, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Ilew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 51339
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (468)..(468)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: (489)..(489)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (520)..(520)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (589)..(589)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-51339

Query Match      12.9%; Score 565.8; DB 17; Length 600;
Best local Similarity 98.6%; Pred. No. 3e-134;
Matches 578; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

Qy      3523 TAATTCACCTAGGGGCTCTCATCTCAACCTTAAGAGAGATTTCTAGAAAACTGGGC 3582
Db      7 TAATTCACCTAGGGGCTCTCATCTCAACCTTAAGAGAGATTTCTAGAAAACTGGGC 66
Qy      3583 CAGATTCTCTTGTCTCATCATATTTAATGAGGAGGCTTCAATTTCTTACTCTTA 3642
Db      67 CAGATTCTCTTGTCTCATCATATTTAATGAGGAGGCTTCAATTTCTTACTCTTA 126
Qy      3643 CCTATGATATTTCTTCTGTAAGTCCCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3702
Db      127 CCTATGATATTTCTTCTGTAAGTCCCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186
Qy      3703 GACTTGTCTTGTATCCCTCAGTTCTTCTTGAATTCAGATGTGTGGGTTCTTAAT 3762
Db      187 GACTTGTCTTGTATCCCTCAGTTCTTCTTGAATTCAGATGTGTGGGTTCTTAAT 246
Qy      3763 TTGGTATGATGATTAAGATTTAAGCATGTGTGTGTGCTTCCAGAGGAGATCCCCAG 3822
Db      247 TTGGTATGATGATTAAGATTTAAGCATGTGTGTGTGCTTCCAGAGGAGATCCCCAG 306
Qy      3823 TTTCTGATGAAGTGAAGTGAAGAGATCAAGAGTGTATCTGAGCAATTAAGTA 3882
Db      307 TTTCTGATGAAGTGAAGTGAAGAGATCAAGAGTGTATCTGAGCAATTAAGTA 366
Qy      3883 GATTCTATTTCTTGTGTCTTCTCTCTCTGAGAGACTCTTAATTTATGTCCCTCTTC 3942
Db      367 GATTCTATTTCTTGTGTCTTCTCTCTCTGAGAGACTCTTAATTTATGTCCCTCTTC 426
Qy      3943 TAGTTAATTTCTTGTATTTGATCTTTGTGAGAGAGTGTGACAGTATTAAGCA 4002
Db      427 TAGTTAATTTCTTGTATTTGATCTTTGTGAGAGAGTGTGACAGTATTAAGCA 486
Qy      4003 AGTTCAAGTGAAGATTAAGTGAAGTGTGAAGTGTGGGGGAAATTAAGCTTAATTTTC 4062
Db      487 AGTTCAAGTGAAGATTAAGTGAAGTGTGAAGTGTGGGGGAAATTAAGCTTAATTTTC 545
Qy      4063 CCTACATGGGATTAACAACACTGTGAATTCATCTTCACTGAAGGCC 4108
Db      546 CCTACATGGGATTAACAACACTGTGAATTCATCTTCACTGAAGGCC 591

RESULT 11
US-10-085-783A-51339
; Sequence 51339, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Ilew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12

```

```

; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 51339
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (468)..(468)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (520)..(520)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (589)..(589)
; OTHER INFORMATION: n is a, c, g, or t
; US-10-085-783A-51339

```

Query Match 12.9%; Score 565.8; DB 17; Length 600;

Best Local Similarity 98.6%; Pred. No. 3e-134; 7; Indels 1; Gaps 1;
Matches 578; Conservative 0; Mismatches

```

QY 3523 TAATTCACAGGAGGCTCTCATCTCACACCTTAAGAGAGATTTCTAGAAAACTGGGC 3582
DB 7 TAATTCACAGGAGGCTCTCATCTCACACCTTAAGAGAGATTTCTAGAAAACTGGGC 66
QY 3583 CAGATTTCTTTGTTCTTCATCATTTTAATGTCGAGGCTGTCAGTTTCTTACTCTTA 3642
DB 67 CAGATTTCTTTGTTCTTCATCATTTTAATGTCGAGGCTGTCAGTTTCTTACTCTTA 126
QY 3643 CCTATGTATATTTCTTGTAACGTGTCACAAAAAGAAAAAGCCCAATCAGTCTCTT 3702
DB 127 CCTATGTATATTTCTTGTAACGTGTCACAAAAAGAAAAAGCCCAATCAGTCTCTT 186
QY 3703 GACTTGTCTTGTGATCCCTCAGTTTCTTGTGATTTGACAGTGTGCGGTTCTTAAT 3762
DB 187 GACTTGTCTTGTGATCCCTCAGTTTCTTGTGATTTGACAGTGTGCGGTTCTTAAT 246
QY 3763 TTGGGTATGATTAAGCAATTTAACCATTGTGTTTGTCCCTACCCAGGGAATCCCCAG 3822
DB 247 TTGGGTATGATTAAGCAATTTAACCATTGTGTTTGTCCCTACCCAGGGAATCCCCAG 306
QY 3823 TTTCGACTTGAAGTGAAGTGAAGAAATCCAGAGTGTATCTGCGCAGATTTAAGTA 3882
DB 307 TTTCGACTTGAAGTGAAGTGAAGAAATCCAGAGTGTATCTGCGCAGATTTAAGTA 366
QY 3883 GATTCTAATTCCTGGTTCCTCCTCCCTGAGAGACCTTATTTATTTGTCCTCTTC 3942
DB 367 GATTCTAATTCCTGGTTCCTCCTCCCTGAGAGACCTTATTTATTTGTCCTCTTC 426
QY 3943 TAGGTTAATTCCTTTGATTTGACTTTGTGAGAAAGAGTTGA CAGTAGATTGCAA 4002
DB 427 TAGGTTAATTCCTTTGATTTGACTTTGTGAGAAAGAGTTGA CAGTAGATTGCAA 486
QY 4003 AATTCCAAATGCAAAATTTACAGTGTGTGAGTGTGGGGGAAAAATTAAGTCTATTTTC 4062
DB 487 AATTCCAAATGCAAAATTTACAGTGTGTGAGTGTGGGGGAAAAATTAAGTCTATTTTC 545
QY 4063 CCTACATGGGATACACACTGTGATTTCAATCTTCACTGAAGGCC 4108
DB 546 CCTACATGGGATACACACTGTGATTTCAATCTTCACTGAAGGCC 591

```

RESULT 12

US-09-917-800A-725

; Sequence 725, Application US/09917800A

; Patent No. US20020119462A1

; GENERAL INFORMATION:

; APPLICANT: Mendrick, Donna

; APPLICANT: Porter, Mark

; APPLICANT: Johnson, Kory

; APPLICANT: Castle, Arthur

; APPLICANT: Blaschoff, Michael

; APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Molecular Toxicology Modeling

; FILE REFERENCE: 44921-5038-US

; CURRENT APPLICATION NUMBER: US/09/917,800A

; PRIOR FILING DATE: 2001-07-31

; PRIOR APPLICATION NUMBER: US 60/222,040

; PRIOR FILING DATE: 2000-07-31

; PRIOR APPLICATION NUMBER: US 60/222,880

; PRIOR FILING DATE: 2000-11-02

; PRIOR APPLICATION NUMBER: US 60/290,029

; PRIOR FILING DATE: 2001-05-11

; PRIOR APPLICATION NUMBER: US 60/290,645

; PRIOR FILING DATE: 2001-05-15

; PRIOR APPLICATION NUMBER: US 60/292,336

; PRIOR FILING DATE: 2001-05-22

; PRIOR APPLICATION NUMBER: US 60/295,798

; PRIOR FILING DATE: 2001-06-06

; PRIOR APPLICATION NUMBER: US 60/297,457

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/298,884

; PRIOR FILING DATE: 2001-06-19

; PRIOR APPLICATION NUMBER: US 60/303,459

; PRIOR FILING DATE: 2001-07-09

; NUMBER OF SEQ ID NOS: 1740

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 725

; LENGTH: 615

; TYPE: DNA

; ORGANISM: Rattus norvegicus

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)..(615)

; OTHER INFORMATION: n = a or c or g or t

; US-09-917-800A-725

Query Match 12.6%; Score 551.6; DB 9; Length 615;

Best Local Similarity 94.6%; Pred. No. 1.4e-130; 30; Indels 3; Gaps 1;
Matches 582; Conservative 0; Mismatches

```

QY 112 CGGCCCTCCGCGGCAATCTGAGCAGCAGTGAATACAGAGCGGCTGCGGCGCTGG 171
DB 1 CGGCCCTCCGCGGCAATCTGAGCAGCAGTGAATACAGAGCGGCTGCGGCGCTGG 60
QY 172 GCGGGCCGAGGACGAGCTGAGGCGGCGGAGGCCCACTTCTCCCGGCAACCGTAAC 231
DB 61 GCGGGCCGAGGACGAGCTGAGGCGGCGGAGGCCCACTTCTCCCGGCAACCGTAAC 120
QY 232 TCAAGAGCGCGGGGCGGCGGCTGCTCTCCCAAGGCGGAGGCCCGGCGCTTCCCGG 291
DB 121 TCAAGAGCGCGGGGCGGCGGCTGCTCTCCCAAGGCGGAGGCCCGGCGCTTCCCGG 180
QY 292 CGGGCT---GCGGCGGCAAGGCGGCGGCTGTTACTCCCGGCGGAGCGGCGGCGGCG 348
DB 181 CGGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 240
QY 349 ACGAGAGAGAGAGCTGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 408
DB 241 ACGAGAGAGAGAGCTGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 409 AAGCGGCAATTGGGGGAGAGCTGCGCGAGCGGAGCGGCGGCTGCGAGCGGCGGCGCAAGT 468

```

Db 301 AAGCGCGATCGGCGGGAGCCAGTCGACGCCGGGCTGGCTGCGAGCCCGCGGCCCAAGT 360

QY 469 ATCAGGCGGTGTGCTGCCCATTCAGACGGGCTCTCTGTTGGCGGGCGGACCAAGAGCTTAAGC 528

Db 361 ATCAGGCGGTGTGCTGCCCATTCAGAGGGGCTCTCTGTTGGCGGGCGGACCAAGAGCTTAAGC 420

QY 529 CTTGGGCTGGGGACAAAGGTTGGGGCGGGCTCCCGCGTGCACCGCTTCGAGACCGGGCGG 588

Db 421 CTTGGGCTGGGGACAAAGGTTGGGGCGGGCTCCCGCGTGCACCGCTTCGAGACCGGGCGG 480

QY 589 GACCCCCACCACTACTCTGTCCCGGGCGGCCACCCCTGGCGGCCACCGGCACCGCGGGA 648

Db 481 GACCCCCACCACTACTCTGTCCCGGGCGGCCACCCCTGGCGGCCACCGGCACCGCGGGA 540

QY 649 CCTTGGCGGGCCAGGCGAGGCGAGATTGAAGAATGAGAGAGGCCCTCTCGGGGGTGGTG 708

Db 541 CCTTGGCGGGCCAGTGAAGGCGAGATTGAAGAATGAGAGAGGCCCTCTCGGGGGTGGTG 600

QY 709 GCGGCTCGGAGGCTT 723

Db 601 GCNGCTCGGAGGCTT 615

```

RESULT 13
US-09-736-457-1554
: Sequence 1554, Application US/09736457
Patent No. US2002016867A1
: GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodges, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedrick, Tom
APPLICANT: Carter, Darick
APPLICANT: Reltter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.4786C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: PaetSeq for Windows Version 3.0
SEQ ID NO 1554
LENGTH: 542,
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(542)
OTHER INFORMATION: n = A,T,C or G
US-09-736-457-1554

```

Query Match	11.7%;	Score 510.2;	DB 9;	Length 542;
Best Local Similarity	98.0%;	Pred. No. 5.4e-120;		
Matches 532; Conservative	4;	Mismatches 4;	Indels 3;	Gaps 2

[illegible]

Db	179	GAGGAATAATGGAATTAAATTCACCTAGAGGGCTCATCTACACCTTAAGGAGAGATT	238
OY	3568	CTAGAAAAAATGAGGAGATTTTCTTGTGTTCTCATCATTTTAAATGAGAGCTGTTC	3627
Db	239	CTAGAAAAAATGAGGAGATTTTCTTGTGTTCTCATCATTTTAAATGAGAGCTGTTC	298
OY	3628	GTTTTCTTACTCTTACTAATGATATTCTTGTAACGTGTCCAAAAAGAAAAAGACC	3687
Db	299	GTTTTCTTACTCTTACTAATGATATTCTTGTAACGTGTCCAAAAAGAAAAAGACC	358
OY	3688	CAATCAATGTCCTTGAACCTTGTCTTGATCCCTGAGTTCTCTGATTTGAGCANGT	3747
Db	359	CAATCAATGTCCTTGAACCTTGTCTTGATCCCTGAGTTCTCTGATTTGAGCANGT	418
OY	3748	GT - GGGGTCCTAATTTGGGATAGATGTGCAAAATTAACCATTTGTTGNCCTTAC	3806
Db	419	GTCGGGGTCCTAATTTGGGATAGATGTGCAAAATTAACCATTTGTTGNCCTTAC	478
OY	3807	CCAGGGAGCTCCACAGTTTCTGAACCTGAAGTGAAGTGAAGATCCACGAGGTGCTATC	3866
Db	479	CCAGGGAGCTCCACAGTTTCTGAACCTGAAGTGAAGTGAAGATCCACGAGGTGCTATC	538
OY	3867	TGG 3869	
Db	539	TTG 541	

```

US-RESULT 14
US-09-902-941-1554
: Sequence 1554, Application US/09902941
: Patent No. US20020172952A1
GENERAL INFORMATION:
APPLICANT: Henderson, Robert A.
APPLICANT: Wang, Tonglong
APPLICANT: Matanabe, Yoshihiro
APPLICANT: Johnson, Jeffrey C.
APPLICANT: Retter, Marc W.
APPLICANT: Marnierakis, Margarita
APPLICANT: Carter, Darick
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: McNabb, Andria
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: OF LUNG CANCER
FILE REFERENCE: 210121.478C17
CURRENT APPLICATION NUMBER: US/09/902,941
CURRENT FILING DATE: 2001-07-10
NUMBER OF SEQ. ID NOS: 2002
SOFTWARE: PatSeq for Windows Version 4.0
SEQ ID NO 1554
LENGTH: 542
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 517_532
OTHER INFORMATION: n = A,T,C or G
US-09-902-941-1554

```

Query Match	11.7%	Score 510.2;	DB 9;	Length 542;
Best Local Similarity	98.0%;	Pred. No. 5.4e-120;		
Matches 532;	Conservative 4;	Mismatches 4;	Indels 3;	Gaps 2;

[illegible]

